

WO02083953

Publication Title:

**METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC
RNA STRUCTURAL MOTIFS**

Abstract:

The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

Data supplied from the esp@cenet database - <http://ep.espacenet.com>

This Patent PDF Generated by Patent Fetcher(TM), a service of Stroke of Color, Inc.

ATTORNEY DOCKET NUMBER: 10589-034-999
SERIAL NUMBER: 10/551,300
REFERENCE: **B04**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083953 A1

(51) International Patent Classification⁷: C12Q 1/68,
C07H 21/02, G01N 27/26

(21) International Application Number: PCT/US02/11757

(22) International Filing Date: 11 April 2002 (11.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/282,965 11 April 2001 (11.04.2001) US

(71) Applicant (for all designated States except US): PTC
THERAPEUTICS, INC. [US/US]; 100 Corporate Court,
Middlesex Business Center, South Plainfield, NJ 07080
(US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): RANDO, Robert
[US/US]; 3 Brown Court, Annandale, NJ 08801 (US).
WELCH, Ellen [US/US]; 33 Hollow Brook Road, Cali-
fon, NJ 07830 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds
LLP, 1155 Avenue of the Americas, New York, NY 10036
(US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Published:

- with international search report
- with amended claims

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/083953 A1

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) Abstract: The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

5 This application claims the benefit of U.S. Provisional Application No.
60/282,965, filed April 11, 2001, which is incorporated herein by reference in its entirety.

1. INTRODUCTION

The present invention relates to a method for screening and identifying test
10 compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-
competitive binding assays are advantageously used to screen libraries of compounds for
those that selectively bind to a preselected target RNA. Binding of target RNA molecules to
a particular test compound is detected using any physical method that measures the altered
physical property of the target RNA bound to a test compound. The methods of the present
15 invention provide a simple, sensitive assay for high-throughput screening of libraries of
compounds to identify pharmaceutical leads.

2. BACKGROUND OF THE INVENTION

Protein-nucleic acid interactions are involved in many cellular functions,
20 including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily
accessible synthetic molecules that can bind with high affinity to specific sequences of
single- or double-stranded nucleic acids have the potential to interfere with these
interactions in a controllable way, making them attractive tools for molecular biology and
medicine. Successful approaches for blocking function of target nucleic acids include using
25 duplex-forming antisense oligonucleotides (Miller, 1996, Progress in Nucl. Acid Res. &
Mol. Biol. 52:261-291; Ojwang & Rando, 1999, Achieving antisense inhibition by
oligodeoxynucleotides containing N₇ modified 2'-deoxyguanosine using tumor necrosis
factor receptor type 1, METHODS: A Companion to Methods in Enzymology 18:244-251)
and peptide nucleic acids ("PNA") (Nielsen, 1999, Current Opinion in Biotechnology
30 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming
anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, RNA 3:850-860;
Aggarwal *et al.*, 1996, Cancer Res. 56:5156-5164; U.S. Patent No. 5,650,316), as well as
pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, Nature 387:202-205; White
et al., 1998, Nature 391:468-471), which are specific for the major and minor grooves of a
35 double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in The Ribosome: Structure, Function & Evolution (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, Biologic activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology" Stein. & Krieg (eds) John Wiley and Sons, New York, pages

335-352). Such structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993; Science 261:1574-1577). The dependence of these functions on the native
5 three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules. Approaches to screening generally involve competitive assays designed to identify
10 compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify
15 compounds that disrupt or interfere with the two components in the assay.

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

20 The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific
25 sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. In particular, the
30 present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the uncomplexed, unlabeled test compounds in the library, and from
35 uncomplexed, labeled target RNA, by a variety of methods, including but not limited to, methods that differentiate changes in the electrophoretic, chromatographic, or thermostable

properties of the complexed target RNA. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. The structure of the test compound attached to the labeled RNA is then determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1 and Section 6.

As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*, ^{32}P , ^{35}S , and ^3H), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin-streptavidin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include a "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamers, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means.

As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an
5 "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a
10 wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may
15 emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in
20 nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric,
25 maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*,
30 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and,
35 particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound, the test compound can be easily identified.

4. DESCRIPTION OF DRAWINGS

FIG. 1. Gel retardation analysis to detect peptide-RNA interactions. In 20 μ l reactions containing increasing concentrations of Tat₄₇₋₅₈ peptide (0.1 μ M, 0.2 μ M, 0.4 μ M, 0.8 μ M, 1.6 μ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. 10 ml of 30% glycerol was added to each sample and applied to a 12% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4°C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated with a phosphorimager. The concentration of peptide added is indicated above each lane.

FIG. 2. Gentamicin interacts with an oligonucleotide corresponding to the 16S rRNA. 20 μ l reactions containing increasing concentrations of gentamicin (1 ng/ml, 10 ng/ml, 100 ng/ml, 1 μ g/ml, 10 μ g/ml, 50 μ g/ml, 500 μ g/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer, heated at 90°C for 2 min and allowed to cool slowly to 24°C. Then 10 μ l of 30% glycerol was added to each sample and the samples were applied to a 13.5% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4°C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated using a phosphorimager. The concentration of gentamicin added is indicated above each lane.

FIG. 3. The presence of 10 pg/ml gentamicin produces a gel mobility shift in the presence of the 16S rRNA oligonucleotide. 20 μ l reactions containing increasing concentrations of gentamicin (100 ng/ml, 10 ng/ml, 1 ng/ml, 100 pg/ml, and 10 pg/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer were treated as described for Figure 2.

FIG. 4. Gentamicin binding to the 16S rRNA oligonucleotide is weak in the absence of MgCl₂. Reaction mixtures containing gentamicin (1 mg/ml, 100 μ g/ml,

10 µg/ml, 1 µg/ml, 0.1 µg/ml, and 10 ng/ml) were treated as described in Figure 2 except that the TKM buffer does not contain MgCl₂.

FIG. 5. Gel retardation analysis to detect peptide-RNA interactions. In reactions
5 containing increasing concentrations of Tat₄₇₋₅₈ peptide (0.1 µM, 0.2 µM, 0.4 µM, 0.8 µM, 1.6 µM) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College,
10 Pennsylvania). The peaks correspond to the amount of free TAR RNA ("TAR") or the Tat-TAR complex ("Tat-TAR"). The concentration of peptide added is indicated below each lane.

5. DETAILED DESCRIPTION OF THE INVENTION

15 The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and
20 a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. Changes in the physical property of the RNA-test compound complex relative to the target RNA or test compound can be measured by methods such as, but not limited to, methods that detect a change in mobility due to a change in mass, change in charge, or a change in thermostability. Such methods
25 include, but are not limited to; electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in
30 solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the unlabeled, uncomplexed test compounds in the library by a variety of methods capable of differentiating changes in the physical properties of the complexed target RNA. The structure of the test compound
35 attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds,

each having an address or identifier, may be deconvoluted, *e.g.*, by cross-referencing the positive sample to an original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR").

Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are deciphered from the input library by methods depending on the type of library that is used. The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

Section 5.1 describes examples of protein-RNA interactions that are important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, *e.g.*, in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 5.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention. Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by

the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

5

5.1. Biologically Important RNA-Host Cell Factor Interactions

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and according affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

15

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

20

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs ("mRNAs"), *e.g.*, protein production, by binding to one or more regulatory elements in the 5' untranslated region, the 3' untranslated region, or the coding region of the mRNA. Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing

30

35

translation of mRNA coding for the protein whose function is decreased, e.g., factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to
5 mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiostatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins
10 and cyclin inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF
15 and thrombopoietin; growth factors such as BNDF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- β , TGF- α and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- α , and TNF- β ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-S transferases, heparanase, HMG CoA
20 synthetase, N-acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, e.g., N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- α reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic
25 hormones, glucagon, gonadotropin, gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotrophin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone, parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system;
30 ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein
35 receptors, opioid pain receptors, substance P receptors, retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue

plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brca1*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 5.1 and Section 6.

The methods of the invention can be used to identify mRNA-binding test compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 5.1 and diseases caused by aberrant expression of the genes disclosed in Example 6. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, *e.g.*, assembly of an *in vivo* protein-RNA or RNA-RNA complex that is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, *e.g.*, as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test

compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with
5 telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human
10 immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target
15 for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and
20 therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some
25 embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-
30 protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets such as mRNAs of INF α , INF γ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required
35 for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF α , INF β and INF γ ; plasmodium falciparum mRNAs; the 16S

ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 6.

5 One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (*e.g.*, from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

10 Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such
15 diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea
20 infection, candida infection, and meningitis.

Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed
25 below.

5.1.1. HIV TAR Element

Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation
30 response region RNA ("TAR RNA"), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-
35 TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA

can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Hamy *et al.*, 1998, Biochemistry 37:5086-5095; Mei *et al.*, 1998, Biochemistry 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

5

5.1.2. Internal Ribosome Entry Site ("IRES")

Internal ribosome entry sites ("IRES") are found in the 5' untranslated regions ("5' UTR") of several mRNAs, and are thought to be involved in the regulation of translational efficiency. When the IRES element is present on an mRNA downstream of a translational stop codon, it directs ribosomal re-entry (Ghattas *et al.*, 1991, Mol. Cell. Biol. 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, approximately 400 nucleotides in length, promotes internal entry of ribosomes independent of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera entero- and rhinoviruses, on the one hand, and cardio- and aphthoviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. The IRES elements of cardio-, entero- and aphthoviruses bind a cellular protein, p57. In the case of cardioviruses, the interaction between a specific stem-loop of the IREs is essential for translation *in vitro*. The IRES elements of entero- and cardioviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* in the absence of any viral gene products, is speculated that IRES-like elements may also occur in specific cellular mRNAs releasing them from cap-dependent translation (Jang *et al.*, 1990, Enzyme 44(1-4):292-309).

30

5.1.3. "Slippery Site"

Programmed, or directed, ribosomal frameshifting, when ribosomes shift from one translation reading frame to another and synthesize two viral proteins from a single viral mRNA, is directed by a unique site in viral mRNAs called the "slippery site." The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the

35

ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to produce a new protein.

5 Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 10 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiaviruses, which are also members of the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craigen & 15 Caskey, 1986, Nature 322:273); bacteriophage genes (Condrón *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, Xenopus, and *Drosophila* ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 20 21:5838-5842).

Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. Compounds that target the RNA elements involved in regulating 25 programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own 30 genome.

5.1.4. Instability Elements

"Instability elements" may be defined as specific sequence elements that 35 promote the recognition of unstable mRNAs by cellular turnover machinery. Instability

elements have been found within mRNA protein coding regions as well as untranslated regions.

Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, producing a form of the *c-myc* mRNA that is five times more stable (*see, e.g.,* Kapstein *et al.*, 1996, J. Biol. Chem. 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 3' UTR adenylate uridylylate rich element ("ARE") that is found in the more labile and weakly oncogenic *c-fos* mRNA (*see, e.g.,* Schiavi *et al.*, 1992, Biochim Biophys Acta. 1114(2-3):95-106). Differences between the benign cervical lesions brought about by nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (*see, e.g.,* Jeon & Lambert, 1995, Proc. Natl. Acad. Sci. USA 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (*see, e.g.,* Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in mast tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. *See, e.g.,* Bickel *et al.*, 1990, J. Immunol. 145(3):840-5.

The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that mis-regulation of the stability of the important modulator TNF α mRNA leads to diseases

such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g.,* Clark, 2000, *Arthritis Res.* 2(3):172-4.

Smooth muscle in the heart is modulated by the β -adrenergic receptor, which in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the β -adrenergic receptor mRNA. *See, e.g.,* Ellis & Frielle, 1999, *Biochem. Biophys. Res. Commun.* 258(3):552-8.

A large number of diseases result from over-expression of collagen. For example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g.,* Lindquist *et al.*, 2000, *Am. J. Physiol. Gastrointest. Liver Physiol.* 279(3):G471-6.

5.1.5. Adenylate Uridylate-rich Elements ("ARE")

Adenylate uridylate-rich elements ("ARE") are found in the 3' untranslated regions ("3' UTR") of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, depending on sequence (Bakheet *et al.*, 2001, *Nucl. Acids Res.* 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

25	Group I Cluster	(AUUUUUUAUUUAUUUAUUUA)	SEQ ID NO: 1
	Group II Cluster	(AUUUUUUAUUUAUUUA) stretch	SEQ ID NO: 2
	Group III Cluster	(WAUUUUUAUUUAUW) stretch	SEQ ID NO: 3
	Group IV Cluster	(WWAUUUUAUUUAWW) stretch	SEQ ID NO: 4
30	Group V Cluster	(WWWWAUUUUAWWWW) stretch	SEQ ID NO: 5

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the 13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, *Nucleic Acids Research*, 29(1): 22-28), in addition to the categories: inflammation, immune response, development/differentiation, using an extensive literature search.

Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- β that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754-4758). Although TNF α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086-21093).

Unlike Group I, Groups II-V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652-7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494-3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144-8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF α , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753-3763; Carballo *et al.*, 2000, Blood 95:1891-1899).

Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

5.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (*e.g.*, fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon GreenTM, rhodamine and derivatives (*e.g.*,

Texas red, tetramethylrhodimine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (*e.g.*, ³H, ¹²⁵I, ³⁵S, ¹⁴C, ³²P, ³³P, *etc.*), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. Useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer-streptavidin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label may be coupled directly or indirectly to a component of the detection assay (*e.g.*, the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (*see, e.g.*, Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, Genetic Engineering 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, Biochemistry, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate.

Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or *t*-butyl amine. The cleavage reaction also removes any phosphate protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or *t*-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, *e.g.*, Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (*e.g.*, a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme

is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

5 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid*
10 *Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and
15 (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP- γ S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

A detectable label should not be incorporated into a target nucleic acid at the
20 specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding
25 region.

After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (*see* Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but
30 are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, *e.g.*, a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

35 In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag₂S, ZnS, CdS, CdTe, Au, or TiO₂. Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

5.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphoric acids and α -amino phosphonic acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinyllogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries.

Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, J. Comb. Chem. 3:477-517; Hall *et al.*, 2001, J. Comb. Chem. 3:125-150; Dolle, 2000, J. Comb. Chem. 2:383-433; Dolle, 1999, J. Comb. Chem. 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses, Vol. 1*, A.W. Czarnik, ed., Wiley Interscience, 2001.

Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, J. Org. Chem. 60:2652; Anderson *et al.*, 1995, J. Org. Chem. 60:2650; Fitch *et al.*, 1994, J. Org. Chem. 59:7955; Look *et al.*, 1994, J. Org. Chem. 49:7588; Metzger *et al.*, 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist *et al.*, 1994, Rapid Commun. Mass Spect. 8:77; Chu *et al.*, 1995, J. Am. Chem. Soc. 117:5419; Brummel *et al.*, 1994, Science 264:399; Stevanovic *et al.*, 1993, Bioorg. Med. Chem. Lett. 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (*see* Lam *et al.*, 1997, Chem. Rev. 97:41-448; Ohlmeyer *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (*see, e.g.,* Nefzi *et al.*, 1997, Chem. Rev. 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (PMBHA) resin (Peptides International, Louisville, KY), polystyrenes (*e.g.,* PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and

aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tübingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden).

In one embodiment, the solid phase support is suitable for *in vivo* use, i.e., it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tübingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestible.

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput. Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

Dynamic combinatorial chemistry uses non-covalent interaction with a target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol

exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in their entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

5.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor (if known) is formed *in vitro*. For example, TK

buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

5 The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA
10 having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

15 The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the types of test compounds in the library, and therefore, the type and amount of the buffer used
20 in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

25 In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. Without
30

being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl₂ stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α -D-glucopyranoside, n-decyl β -D-maltopyranoside, n-dodecyl β -D-maltoside, n-octyl β -D-glucopyranoside, sorbitan esters, n-tetradecyl β -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxypolyethoxyethanol (NP-40), and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, e.g., bovine serum albumin ("BSA"), before contacting with to the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is

substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using an electrophoretic technique (see Section 5.5.1), or any of the methods disclosed in Section 5.5 *infra*. In another embodiment, the complexed target nucleic acid does not need to be separated from the free target nucleic acid if a technique (*i.e.*, spectrometry) that differentiates between bound and unbound target nucleic acids is used.

The methods for identifying small molecules bound to labeled nucleic acid will vary with the type of label on the target nucleic acid. For example, if a target RNA is labeled with a visible or fluorescent dye, the target RNA complexes are preferably identified using a chromatographic technique that separates bound from free target by an electrophoretic or size differential technique using individual reactions. The reactions corresponding to changes in the migration of the complexed RNA can be cross-referenced to the small molecule compound(s) added to said reaction. Alternatively, complexed target RNA can be screened *en masse* and then separated from free target RNA using an electrophoretic or size differential technique, the resultant complexed target is then analyzed using a mass spectrometric technique. In this fashion the bound small molecule can be identified on the basis of its molecular weight. In this reaction *a priori* knowledge of the exact molecular weights of all compounds within the library is known. In another embodiment, the test compounds bound to the target nucleic acid may not require separation from the unbound target nucleic acid if a technique such as, but not limited to, spectrometry is used.

5.5. Separation Methods for Screening Test Compounds

Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids. Methods that can be utilized for the physical separation of complexed target RNA from unbound target RNA include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography,

and nanoparticle aggregation.

5.5.1. Electrophoresis

5 Methods for separation of the complex of a target RNA bound to a test compound from the unbound RNA comprises any method of electrophoretic separation, including but not limited to, denaturing and non-denaturing polyacrylamide gel electrophoresis, urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, agarose gel electrophoresis, and capillary electrophoresis.

10 In a preferred embodiment, an automated electrophoretic system comprising a capillary cartridge having a plurality of capillary tubes is used for high-throughput screening of test compounds bound to target RNA. Such an apparatus for performing automated capillary gel electrophoresis is disclosed in U.S. Patent Nos. 5,885,430; 5,916,428; 6,027,627; and 6,063,251, the disclosures of which are incorporated by
15 reference in their entirety.

 The device disclosed in U.S. Patent No. 5,885,430, which is incorporated by reference in its entirety, allows one to simultaneously introduce samples into a plurality of capillary tubes directly from microtiter trays having a standard size. U.S. Patent No. 5,885,430 discloses a disposable capillary cartridge which can be cleaned between
20 electrophoresis runs, the cartridge having a plurality of capillary tubes. A first end of each capillary tube is retained in a mounting plate, the first ends collectively forming an array in the mounting plate. The spacing between the first ends corresponds to the spacing between the centers of the wells of a microtiter tray having a standard size. Thus, the first ends of the capillary tubes can simultaneously be dipped into the samples present in the tray's wells.
25 The cartridge is provided with a second mounting plate in which the second ends of the capillary tubes are retained. The second ends of the capillary tubes are arranged in an array which corresponds to the wells in the microtiter tray, which allows for each capillary tube to be isolated from its neighbors and therefore free from cross-contamination, as each end is dipped into an individual well.

30 Plate holes may be provided in each mounting plate and the capillary tubes inserted through these plate holes. In such a case, the plate holes are sealed airtight so that the side of the mounting plate having the exposed capillary ends can be pressurized. Application of a positive pressure in the vicinity of the capillary openings in this mounting plate allows for the introduction of air and fluids during electrophoretic operations and also
35 can be used to force out gel and other materials from the capillary tubes during

reconditioning. The capillary tubes may be protected from damage using a needle comprising a cannula and/or plastic tubes, and the like when they are placed in these plate holes. When metallic cannula or the like are used, they can serve as electrical contacts for current flow during electrophoresis. In the presence of a second mounting plate, the second mounting plate is provided with plate holes through which the second ends of the capillary tubes project. In this instance, the second mounting plate serves as a pressure containment member of a pressure cell and the second ends of the capillary tubes communicate with an internal cavity of the pressure cell. The pressure cell is also formed with an inlet and an outlet. Gels, buffer solutions, cleaning agents, and the like may be introduced into the internal cavity through the inlet, and each of these can simultaneously enter the second ends of the capillaries.

In another preferred embodiment, the automated electrophoretic system can comprise a chip system consisting of complex designs of interconnected channels that perform and analyze enzyme reactions using part of a channel design as a tiny, continuously operating electrophoresis material, where reactions with one sample are going on in one area of the chip while electrophoretic separation of the products of another sample is taking place in a different part of the chip. Such a system is disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties.

The system disclosed in U.S. Patent No. 5,699,157, which is hereby incorporated by reference in its entirety, provides for a microfluidic system for high-speed electrophoretic analysis of subject materials for applications in the fields of chemistry, biochemistry, biotechnology, molecular biology and numerous other areas. The system has a channel in a substrate, a light source and a photoreceptor. The channel holds subject materials in solution in an electric field so that the materials move through the channel and separate into bands according to species. The light source excites fluorescent light in the species bands and the photoreceptor is arranged to receive the fluorescent light from the bands. The system further has a means for masking the channel so that the photoreceptor can receive the fluorescent light only at periodically spaced regions along the channel. The system also has an unit connected to analyze the modulation frequencies of light intensity received by the photoreceptor so that velocities of the bands along the channel are determined, which allows the materials to be analyzed.

The system disclosed in U.S. Patent No. 5,699,157 also provides for a

method of performing high-speed electrophoretic analysis of subject materials, which comprises the steps of holding the subject materials in solution in a channel of a microfluidic system; subjecting the materials to an electric field so that the subject materials move through the channel and separate into species bands; directing light toward the channel; receiving light from periodically spaced regions along the channel simultaneously; and analyzing the frequencies of light intensity of the received light so that velocities of the bands along the channel can be determined for analysis of said materials. The determination of the velocity of a species band determines the electrophoretic mobility of the species and its identification.

U.S. Patent No. 5,842,787, which is hereby incorporated by reference in its entirety, is generally directed to devices and systems employ channels having, at least in part, depths that are varied over those which have been previously described (such as the device disclosed in U.S. Patent No. 5,699,157), wherein said channel depths provide numerous beneficial and unexpected results such as but not limited to, a reduction in sample perturbation, reduced non-specific sample mixture by diffusion, and increased resolution.

In another embodiment, the electrophoretic method of separation comprises polyacrylamide gel electrophoresis. In a preferred embodiment, the polyacrylamide gel electrophoresis is non-denaturing, so as to differentiate the mobilities of the target RNA bound to a test compound from free target RNA. If the polyacrylamide gel electrophoresis is denaturing, then the target RNA:test compound complex must be cross-linked prior to electrophoresis to prevent the disassociation of the target RNA from the test compound during electrophoresis. Such techniques are well known to one of skill in the art.

In one embodiment of the method, the binding of test compounds to target nucleic acid can be detected, preferably in an automated fashion, by gel electrophoretic analysis of interference footprinting. RNA can be degraded at specific base sites by enzymatic methods such as ribonucleases A, U₂, CL₃, T₁, Phy M, and *B. cereus* or chemical methods such as diethylpyrocarbonate, sodium hydroxide, hydrazine, piperidine formate, dimethyl sulfate, [2,12-dimethyl-3,7,11,17-tetraazacyclo[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato] nickel(II) (NiCR), cobalt(II)chloride, or iron(II) ethylenediaminetetraacetate (Fe-EDTA) as described for example in Zheng *et al.*, 1999, *Biochem.* 37:2207-2214; Latham & Cech, 1989, *Science* 245:276-282; and Sambrook *et al.*, 2001, in *Molecular Cloning: A Laboratory Manual*, pp 12.61-12.73, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties. The

specific pattern of cleavage sites is determined by the accessibility of particular bases to the reagent employed to initiate cleavage and, as such, is therefore is determined by the three-dimensional structure of the RNA.

5 The interaction of small molecules with a target nucleic acid can change the accessibility of bases to these cleavage reagents both by causing conformational changes in the target nucleic acid or by covering a base at the binding interface. When a test compound binds to the nucleic acid and changes the accessibility of bases to cleavage reagents, the observed cleavage pattern will change. This method can be used to identify and characterize the binding of small molecules to RNA as described, for example, by
10 Prudent *et al.*, 1995, J. Am. Chem. Soc. 117:10145-10146 and Mei *et al.*, 1998, Biochem. 37:14204-14212.

 In the preferred embodiment of this technique, the detectably labeled target nucleic acid is incubated with an individual test compound and then subjected to treatment
15 with a cleavage reagent, either enzymatic or chemical. The reaction mixture can be preferably be examined directly, or treated further to isolate and concentrate the nucleic acid. The fragments produced are separated by electrophoresis and the pattern of cleavage can be compared to a cleavage reaction performed in the absence of test compound. A change in the cleavage pattern directly indicates that the test compound binds to the target
20 nucleic acid. Multiple test compounds can be examined both in parallel and serially.

 Other embodiments of electrophoretic separation include, but are not limited to urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, and agarose gel electrophoresis.

25

5.5.2. Fluorescence Spectroscopy

 In a preferred embodiment, fluorescence polarization spectroscopy, an optical detection method that can differentiate the proportion of a fluorescent molecule that is either bound or unbound in solution (*e.g.*, the labeled target nucleic acid of the present invention), can be used to read reaction results without electrophoretic separation of the
30 samples. Fluorescence polarization spectroscopy can be used to read the reaction results in the chip system disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties. The application of fluorescence
35

polarization spectroscopy to the chip system disclosed in the U.S. Patents listed *supra* is fast, efficient, and well-adapted for high-throughput screening.

In another embodiment, a compound that has an affinity for the target nucleic acid of interest can be labeled with a fluorophore to screen for test compounds that bind to the target nucleic acid. For example, a pyrene-containing aminoglycoside analog was used to accurately monitor antagonist binding to a prokaryotic 16S rRNA A site (which comprises the natural target for aminoglycoside antibiotics) in a screen using a fluorescence quenching technique in a 96-well plate format (Hamasaki & Rando, 1998, Anal. Biochem. 261(2):183-90).

In another embodiment, fluorescence resonance energy transfer (FRET) can be used to screen for test compounds that bind to the target nucleic acid. FRET, a characteristic change in fluorescence, occurs when two fluorophores with overlapping emission and excitation wavelength bands are held together in close proximity, such as by a binding event. In the preferred embodiment, the fluorophore on the target nucleic acid and the fluorophore on the test compounds will have overlapping excitation and emission spectra such that one fluorophore (the donor) transfers its emission energy to excite the other fluorophore (the acceptor). The acceptor preferably emits light of a different wavelength upon relaxing to the ground state, or relaxes non-radiatively to quench fluorescence. FRET is very sensitive to the distance between the two fluorophores, and allows measurement of molecular distances less than 10 nm. For example, U.S. Patent 6,337,183 to Arenas *et al.*, which is incorporated by reference in its entirety, describes a screen for compounds that bind RNA that uses FRET to measure the effect of test compounds on the stability of a target RNA molecule where the target RNA is labeled with both fluorescent acceptor and donor molecules and the distance between the two fluorophores as determined by FRET provides a measure of the folded structure of the RNA. Matsumoto *et al.* (2000, Bioorg. Med. Chem. Lett. 10:1857-1861) describe a system where a peptide that binds to HIV-1 TAR RNA is labeled on one end with a fluorescein fluorophore and a tetramethylrhodamine on the other end. The conformational change of the peptide upon binding to the RNA provided a FRET signal to screen for compounds that bound to the TAR RNA.

In the preferred embodiment, both the target nucleic acid and a compound that has an affinity for the target nucleic acid of interest are labeled with fluorophores with overlapping emission and excitation spectra (donor and acceptor), including but not limited to fluorescein and derivatives, rhodamine and derivatives, cyanine dyes and derivatives, bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, pyrene, nanoparticles, or

non-fluorescent quenching molecules. Binding of a labeled test compound to the target nucleic acid can be identified by the change in observable fluorescence as a result of FRET.

If the target nucleic acid is labeled with the donor fluorophore, then the test compounds is labeled with the acceptor fluorophore. Conversely, if the target nucleic acid is labeled with the acceptor fluorophore, then the test compounds is labeled with the donor fluorophore. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. The fluorophore on the target nucleic acid must be in close proximity to the binding site of the test compounds, but should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site.

In yet another embodiment, homogeneous time-resolved fluorescence ("HTRF") techniques based on time-resolved energy transfer from lanthanide ion complexes to a suitable acceptor species can be adapted for high-throughput screening for inhibitors of RNA-protein complexes (Hemmilä, 1999, J. Biomol. Screening 4:303-307; Mathis, 1999, J. Biomol. Screening 4:309-313). HTRF is similar to fluorescence resonance energy transfer using conventional organic dye pairs, but has several advantages, such as increased sensitivity and efficiency, and background elimination (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356).

Fluorescence spectroscopy has traditionally been used to characterize DNA-protein and protein-protein interactions, but fluorescence spectroscopy has not been widely used to characterize RNA-protein interactions because of an interfering absorption of RNA nucleotides with the intrinsic tryptophan fluorescence of proteins (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356.). However, fluorescence spectroscopy has been used in studying the single tryptophan residue within the arginine-rich RNA-binding domain of Rev protein and its interaction with the RRE in a time-resolved fluorescence study (Kwon & Carson, 1998, Anal. Biochem. 264:133-140). Thus, in this invention, fluorescence spectroscopy is less preferred if the test compounds or peptides or proteins possess intrinsic tryptophan fluorescence. However, fluorescence spectroscopy can be used for test compounds that do not possess intrinsic fluorescence.

5.5.3. Surface Plasmon Resonance ("SPR")

Surface plasmon resonance (SPR) can be used for determining kinetic rate constants and equilibrium constants for macromolecular interactions by following the

association project in "real time" (Schuck, 1997, *Annu. Rev. Biophys. Biomol. Struct.* 26:541-566).

The principle of SPR is summarized by Xavier *et al.* (*Trends Biotechnol.*, 2000, 18(8):349-356) as follows. Total internal reflection occurs at the boundary between two substances of different refractive index. The incident light's electromagnetic field penetrates beyond the interface as an evanescent wave, which extends a few hundred nanometers beyond the surface into the medium. Insertion of a thin gold foil at the interface produced SPR owing to the absorption of the energy from the evanescent wave by free electron clouds of the metal (plasmons). As a result of this absorbance, there is a drop in the intensity of the reflected light at a particular angle of incidence. The evanescent wave profile depends exquisitely on the refractive index of the medium it probes. Thus, the angle at which absorption occurs is very sensitive to the refractive changes in the external medium. All proteins and nucleic acids are known to change the refractive index of water by a similar amount per unit mass, irrespective of their amino acid or nucleotide composition (the refractive index change is different for proteins and nucleic acids). When the protein or nucleic acid content of the layer at the sensor changes, the refractive index also changes. Typically, one member of a complex is immobilized in a dextran layer and then the other member is introduced into the solution, either in a flow cell (Biacore AB, Uppsala, Sweden) or a stirred cuvette (Affinity Sensors, Santa Fe, New Mexico). It has been determined that there is a linear correlation between the surface concentration of protein or nucleic acid and the shift in resonance angle, which can be used to quantitate kinetic rate constants and/or the equilibrium constants.

In the present invention, the target RNA may be immobilized to the sensor surface through a streptavidin-biotin linkage, the linkage of which is disclosed by Crouch *et al.* (*Methods Mol. Biol.*, 1999, 118:143-160). The RNA is biotinylated either during synthesis or post-synthetically via the conversion of the 3' terminal ribonucleoside of the RNA into a reactive free amino group or using a T7 polymerase incorporated guanosine monophosphorothioate at the 5' end. SPR has been used to determine the stoichiometry and affinity of the interaction between the HIV Rev protein and the RRE (Van Ryk & Venkatesan, 1999, *J. Biol. Chem.* 274:17452-17463) and the aminoglycoside antibiotics with RRE and a model RNA derived from the 16S ribosomal A site, respectively (Hendrix *et al.*, 1997, *J. Am. Chem. Soc.* 119:3641-3648; Wong *et al.*, 1998, *Chem. Biol.* 5:397-406).

In one embodiment of the present invention, the target nucleic acid can be immobilized to a sensor surface (*e.g.*, by a streptavidin-biotin linkage) and SPR can be used

to (a) determine whether the target RNA binds a test compound and (b) further characterize the binding of the target nucleic acids of the present invention to a test compound.

5.5.4. Mass Spectrometry

5

An automated method for analyzing mass spectrometer data which can analyze complex mixtures containing many thousands of components and can correct for background noise, multiply charged peaks and atomic isotope peaks is described in U.S. Patent No. 6,147,344, which is hereby incorporated by reference in its entirety. The system disclosed in U.S. Patent No. 6,147,344 is a method for analyzing mass spectrometer data in which a control sample measurement is performed providing a background noise check. The peak height and width values at each m/z ratio as a function of time are stored in a memory. A mass spectrometer operation on a material to be analyzed is performed and the peak height and width values at each m/z ratio versus time are stored in a second memory location. The mass spectrometer operation on the material to be analyzed is repeated a fixed number of times and the stored control sample values at each m/z ratio level at each time increment are subtracted from each corresponding one from the operational runs, thus producing a difference value at each mass ratio for each of the multiple runs at each time increment. If the MS value minus the background noise does not exceed a preset value, the m/z ratio data point is not recorded, thus eliminating background noise, chemical noise and false positive peaks from the mass spectrometer data. The stored data for each of the multiple runs is then compared to a predetermined value at each m/z ratio and the resultant series of peaks, which are now determined to be above the background, is stored in the m/z points in which the peaks are of significance.

25

One possibility for the utilization of mass spectrometry in high throughput screening is the integration of SPR with mass spectrometry. Approaches that have been tried are direct analysis of the analyte retained on the sensor chip and mass spectrometry with the eluted analyte (Sonksen *et al.*, 1998, *Anal. Chem.* 70:2731-2736; Nelson & Krone, 1999, *J. Mol. Recog.* 12:77-93). Further developments, especially in the interfacing of the sensor chip with the mass spectrometer and in reusing the sensor chip, are required to make SPR combined with mass spectroscopy a high-throughput method for biomolecular interaction analysis and the screening of targets for small molecule inhibitors (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

30

In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by any of the mass spectrometry processed described *supra*. Furthermore, mass spectrometry can also be used to elucidate

35

the structure of the test compound.

5.5.5. Scintillation Proximity Assay ("SPA")

5 Scintillation Proximity Assay ("SPA") is a method that can be used for screening small molecules that bind to the target RNAs. SPA would involve radiolabeling either the target RNA or the test compound and then quantitating its binding to the other member to a bead or a surface impregnated with a scintillant (Cook, 1996, Drug Discov. Today 1:287-294). Currently, fluorescence-based techniques are preferred for high-throughput screening (Pope *et al.*, 1999, Drug Discov. Today 4:350-362).

10 Screening for small molecules that inhibit Tat peptide:TAR RNA interaction has been performed with SPA, and inhibitors of the interaction were isolated and characterized (Mei *et al.*, 1997, Bioorg. Med. Chem. 5:1173-1184; Mei *et al.*, 1998, Biochemistry 37:14204-14212). A similar approach can be used to identify small molecules that directly bind to a preselected target RNA element in accordance with the invention can be utilized.

15 SPA can be adapted to high throughput screening by the availability of microplates, wherein the scintillant is directly incorporated into the plastic of the microtiter wells (Nakayama *et al.*, 1998, J. Biomol. Screening 3:43-48). Thus, one embodiment of the present invention comprises (a) labeling of the target nucleic acid with a radioactive or
20 fluorescent label; (b) contacted the labeled nucleic acid with test compounds, wherein each test compound is in a microtiter well coated with scintillant and is tethered to the microtiter well; and (c) identifying and quantifying the test compounds bound to the target nucleic acid with SPA, wherein the test compound is identified by virtue of its location in the microplate.

25

5.5.6. Structure-Activity Relationships ("SAR") by NMR Spectroscopy

NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects, and NMR-based approaches have been used in
30 the identification of small molecule binders of protein drug targets (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). The determination of structure-activity relationships ("SAR") by NMR is the first method for NMR described in which small molecules that bind adjacent subsites are identified by two-dimensional ^1H - ^{15}N spectra of the target protein (Shuker *et al.*, 1996, Science 274:1531-1534). The signal from the bound molecule is
35 monitored by employing line broadening, transferred NOEs and pulsed field gradient

diffusion measurements (Moore, 1999, Curr. Opin. Biotechnol. 10:54-58). A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, 1999, Chem. Biol. 6:755-769).

5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by SAR by NMR. Furthermore, SAR by NMR can also be used to elucidate the structure of the test compound.

5.5.7. Size Exclusion Chromatography

10 In another embodiment of the present invention, size-exclusion chromatography is used to purify test compounds that are bound to a target nucleic acid from a complex mixture of compounds. Size-exclusion chromatography separates molecules based on their size and uses gel-based media comprised of beads with specific size distributions. When applied to a column, this media settles into a tightly packed matrix and
15 forms a complex array of pores. Separation is accomplished by the inclusion or exclusion of molecules by these pores based on molecular size. Small molecules are included into the pores and, consequently, their migration through the matrix is retarded due to the added distance they must travel before elution. Large molecules are excluded from the pores and migrate with the void volume when applied to the matrix. In the present invention, a target
20 nucleic acid is incubated with a mixture of test compounds while free in solution and allowed to reach equilibrium. When applied to a size exclusion column, test compounds free in solution are retained by the column, and test compounds bound to the target nucleic acid are passed through the column. In a preferred embodiment, spin columns commonly used for "desalting" of nucleic acids will be employed to separate bound from unbound test
25 compounds (*e.g.*, Bio-Spin columns manufactured by BIO-RAD). In another embodiment, the size exclusion matrix is packed into multiwell plates to allow high throughput separation of mixtures (*e.g.*, PLASMID 96-well SEC plates manufactured by Millipore).

5.5.8. Affinity Chromatography

30 In one embodiment of the present invention, affinity capture is used to purify test compounds that are bound to a target nucleic acid labeled with an affinity tag from a complex mixture of compounds. To accomplish this, a target nucleic acid labeled with an affinity tag is incubated with a mixture of test compounds while free in solution and then captured to a solid support once equilibrium has been established; alternatively, target
35 nucleic acids labeled with an affinity tag can be captured to a solid support first and then allowed to reach equilibrium with a mixture of test compounds.

The solid support is typically comprised of, but not limited to, cross-linked agarose beads that are coupled with a ligand for the affinity tag. Alternatively, the solid support may be a glass, silicon, metal, or carbon, plastic (polystyrene, polypropylene) surface with or without a self-assembled monolayer (SAM) either with a covalently
5 attached ligand for the affinity tag, or with inherent affinity for the tag on the target nucleic acid.

Once the complex between the target nucleic acid and test compound has reached equilibrium and has been captured, one skilled in the art will appreciate that the retention of bound compounds and removal of unbound compounds is facilitated by
10 washing the solid support with large excesses of binding reaction buffer. Furthermore, retention of high affinity compounds and removal of low affinity compounds can be accomplished by a number of means that increase the stringency of washing; these means include, but are not limited to, increasing the number and duration of washes, raising the salt concentration of the wash buffer, addition of detergent or surfactant to the wash buffer,
15 and addition of non-specific competitor to the wash buffer.

In one embodiment, the test compounds themselves are detectably labeled with fluorescent dyes, radioactive isotopes, or nanoparticles. When the test compounds are applied to the captured target nucleic acid in a spatially addressed fashion (*e.g.*, in separate wells of a 96-well microplate), binding between the test compounds and the target nucleic
20 acid can be determined by the presence of the detectable label on the test compound using fluorescence.

Following the removal of unbound compounds, bound compounds with high affinity for the target nucleic acid can be eluted from the immobilized target nucleic acids and analyzed. The elution of test compounds can be accomplished by any means that break
25 the non-covalent interactions between the target nucleic acid and compound. Means for elution include, but are not limited to, changing the pH, changing the salt concentration, the application of organic solvents, and the application of molecules that compete with the bound ligand. In a preferred embodiment, the means employed for elution will release the compound from the target RNA, but will not effect the interaction between the affinity tag
30 and the solid support, thereby achieving selective elution of test compound. Moreover, a preferred embodiment will employ an elution buffer that is volatile to allow for subsequent concentration by lyophilization of the eluted compound (*e.g.*, 0 M to 5 M ammonium acetate).

35

5.5.9. Nanoparticle Aggregation

In one embodiment of the present invention, both the target nucleic acid and the test compounds are labeled with nanoparticles. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag₂S, ZnS, CdS, CdTe, Au, or TiO₂. Methods for the attachment of nucleic acids and small molecules to nanoparticles are well known to one of skill in the art (reviewed in Niemeyer, 2001, Angew. Chem. Int. Ed. 40:4129-4158. The references cited therein are hereby incorporated by reference in their entireties). In particular, if multiple copies of the target nucleic acid are attached to a single nanoparticle and multiple copies of a test compound are attached to another nanoparticle, then interaction between the test compound and target nucleic acid will induce aggregation of nanoparticles as described, for example, by Mitchel *et al.* 1999, J. Am. Chem. Soc. 121:8122-8123. The aggregate can be detected by changes in absorbance or fluorescence spectra and physically separated from the unbound components through filtration or centrifugation.

5.6. Methods for Identifying or Characterizing the Test Compounds Bound to the Target Nucleic Acids

If the library comprises arrays or microarrays of test compounds, wherein each test compound has an address or identifier, the test compound can be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the test compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the de novo characterization of test compounds bound to the target.

5.6.1. Mass Spectrometry

Mass spectrometry (*e.g.*, electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound. Thus, one example of mass spectroscopy is that separation of a bound and unbound complex and test compound structure elucidation can be carried out in a single step.

MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, J. Biol. Chem. 272:32084-32091).

5 However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, Anal. Chem. 69:5130-5135).

Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, Anal. Chem. 71:3436-3440; Griffey *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

20 An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5.6.2. NMR Spectroscopy

As described above, NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

35 Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure

of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5

5.6.3. Vibrational Spectroscopy

Vibrational spectroscopy (*e.g.* infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

10

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

20

25

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, must are observable only with one or

35

the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

5 Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

10 Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby
15 incorporated by reference in its entirety.

In the preferred embodiment, test compounds can be identified by matching the IR or Raman spectra of a test compound to a dataset of vibrational (IR or Raman) spectra previously acquired for each compound in the combinatorial library. By this
20 method, the spectra of compounds with known structure are recorded so that comparison with these spectra can identify compounds again when isolated from RNA binding experiments.

5.7. Secondary Biological Screens

25 The test compounds identified in the binding assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the
30 lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene,
35 such as, but not limited to, β -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and β -

lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure ("TRAP") has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, Cell 87(1):115-25; Huang & Schreiber, 1997, Proc. Natl. Acad. Sci. USA 94:13396-13401).

In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling (cf. Lamothe, *et al.* 1997, J. Med. Chem. 40: 3542; Mottola *et al.* 1996, J. Med. Chem. 39: 285; Beusen *et al.* 1995, Biopolymers 36: 181; P. Fossa *et al.* 1998, Comput. Aided Mol. Des. 12: 361), QSAR development (cf. Siddiqui *et al.* 1999, J. Med. Chem. 42: 4122; Barreca *et al.* 1999 Bioorg. Med. Chem. 7: 2283; Kroemer *et al.* 1995, J. Med. Chem. 38: 4917; Schaal *et al.* 2001, J. Med. Chem. 44: 155; Buolamwini & Assefa 2002, J. Mol. Chem. 45: 84), Virtual docking and screening/scoring (cf. Anzini *et al.* 2001, J. Med. Chem. 44: 1134; Faaland *et al.* 2000, Biochem. Cell. Biol. 78: 415; Silvestri *et al.* 2000, Bioorg. Med. Chem. 8: 2305; J. Lee *et al.* 2001, Bioorg. Med. Chem. 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described

by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 J. Mol. Biol. 288: 911-940); RNAMotif (Macke *et al.* 2001, Nucleic Acids Res. 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, Monatsh. Chem. 125: 167-188).

Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinaly & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew *et al.*, 1989, J. Am. Chem. Soc. 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (*e.g.*, Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (*e.g.*, Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (*e.g.*, QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" software (*e.g.* RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (*e.g.*, Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, Adv. Drug Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is

associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

5 In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a
10 discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction
15 *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family
20 history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be
25 administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules,
30 capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral,
35 sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to

the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

5 In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or
10 fibers.

In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a
15 reservoir, such as an Ommaya reservoir.

Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with
20 traditional binders and vehicles such as triglycerides.

In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).
25

In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be
30 used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507 Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and*
35 *Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science*

228:190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

5 Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

10 In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such
15 pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to
20 a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk,
25 silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

 Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-
30 release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see *e.g.*, U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed.,
35 Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The
5 precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound
10 or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer
15 to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01
20 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50
25 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200
30 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

35 The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use

in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

5 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphonic acids and α -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as
10 phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

15 6. EXAMPLE: THERAPEUTIC TARGETS

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

20 6.1. Tumor Necrosis Factor Alpha ("TNF- α ")

GenBank Accession # X01394:

1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca
61 gacgccacat ccctgacaa gctgccaggc aggttctctt cctctcacat actgacccac
121 ggctccaccc tctctccct ggaaggaca ccatgagcac tgaagcatg atccgggacg
25 181 tggagctggc cgaggaggcg ctcccaaga agacaggggg gccccagggc tccaggcggt
241 gcttgttctt cagcctcttc tccttcctga tcgtggcagg cgccaccacg ctcttctgcc
301 tgctgcactt tggagtgate ggccccaga gggaagagtt cccaggggac ctctctctaa
361 tcagccctct ggcccaggca gtcagatcat ctctcgaac cccagtgac aagcctgtag
421 cccatgttgt agcaaaccct caagctgagg ggcagctcca gtggctgaac cgccggggcca
30 481 atgccctctt ggccaatggc gtggagctga gagataacca gctggtggtg ccatcagagg
541 gcctgtacct catctactcc caggtcctct tcaagggcca aggtgcctcc tccaccatg
601 tgctctcac ccacaccatc agccgcacg cgtctccta ccagaccaag gtcaacctcc
661 tctctgcat caagagcccc tgccagaggg agacccaga gggggctgag gccaaagccct
721 ggtatgagcc catctatctg ggagggtgt tccagctgga gaagggtgac cgactcagcg
35 781 ctgagatcaa tcggcccgac tatctcgact ttgccgagtc tgggcaggtc tactttggga

841 tcattgccct gtgaggagga cgaacatcca accttccaa acgcctcccc tgccccaatc
 901 cctttattac cccctcctc agacaccctc aacctcttct ggctcaaaaa gagaattggg
 961 ggcttagggc cggaacccaa gcttagaact ttaagcaaca agaccaccac ttgaaacct
 5 1021 gggattcagg aatgtgtggc ctgcacagtg aattgctggc aaccactaag aattcaaact
 1081 ggggcctcca gaactcactg gggcctacag ctttgatccc tgacatctgg aatctggaga
 1141 ccaggagacc ttggttctg gccagaatgc tgcaggactt gagaagacct cacctagaaa
 1201 ttgacacaag tggaccttag gccttctct ctccagatgt ttccagactt ccttgagaca
 1261 cggagcccag ccctcccat ggagccagct ccctctattt atgtttgcac ttgtgattat
 10 1321 ttattattta ttattattt attatttacc agatgaatgt atttattgg gagaccgggg
 1381 taccctgggg gacccaatgt aggagctgcc ttggctcaga catgtttcc gtgaaaacgg
 1441 agctgaacaa taggctgttc ccatgtagcc cctggcctc tgtgccttct ttgattatg
 1501 tttttaaaa tatttatctg attaatgtgt ctaaacaatg ctgatttggg gaccaactgt
 1561 cactcattgc tgagcctctg ctccccaggg gagttgtgtc tgtaatgcc ctactattca
 15 1621 gtggcgagaa ataaagttg ctt (SEQ ID NO: 6)

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 152
- (2) 3' Untranslated Region - nts 852 - 1643

20 Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUUUUUUUUUUUUUUUUU 3' (SEQ ID NO: 1)

25 6.2. Granulocyte-macrophage Colony Stimulating Factor ("GM-CSF")

GenBank Accession # NM_000758:

1 gctggaggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg
 61 caccgccccg ctgcccagc cccagcacgc agccctggga gcatgtgaat gccatccagg
 121 aggccccggc tctctgaac ctgagtagag aactgctgc tgagatgaat gaaacagtag
 30 181 aagtcattc agaatgttt gacctcagg agccgacctg cctacagacc cgcctggagc
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg
 301 ccagccacta caagcagcac tgccctcaa ccccggaac ttctgtgca acccagacta
 361 tcaccttga aagttcaaa gagaacctga aggactttct gcttgcac cccttgact
 421 gctgggagcc agtcaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatggctac ttggaggagc caaggggtgg
 35 541 gccacagcca tgggtggagt ggctggacc tgccctgggc cactgacc ctgatacagg

601 catggcagaa gaatgggaat atttatact gacagaaatc agtaatatTT atatattat
 661 attttaaaa tatttattta tttatttatt taagttcata ttccatattt attcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 7)

5

GenBank Accession # XM_003751:

1 tctggaggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg
 61 caccgccccg ctcgcccagc cccagcacgc agccctggga gcatgtgaat gccatccagg
 121 agggccggcg tctcctgaac ctgagtagag aactgctgc tgagatgaat gaaacagtag
 10 181 aagtcatttc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg
 301 ccagccacta caagcagcac tgccctccaa ccccggaac ttctgtgca acccagacta
 361 tcaccttga aagttcaaa gagaacctga aggactttct gctgtcatc cccttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 15 481 tctctcatga aacaagagct agaaactcag gatggtcatc ttggaggagc caaggggtgg
 541 gccacagcca tgggtggagt ggctggacc tgccctgggc cacactgacc ctgatacagg
 601 catggcagaa gaatgggaat atttatact gacagaaatc agtaatatTT atatattat
 661 attttaaaa tatttattta tttatttatt taagttcata ttccatattt attcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 8)

20

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 32
- (2) 3' Untranslated Region - nts 468 - 789

25

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUAUUUAUUUAUUUAUUUA 3' (SEQ ID NO: 1)

6.3. Interleukin 2 ("IL-2")

30

GenBank Accession # U25676:

1 atcactctct ttaactacta ctacattaa cctcaactcc tgccacaatg tacaggatgc
 61 aactcctgtc ttgcattgca ctaattcttg cactgtcac aaacagtgc cctactcaa
 121 gtgcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggatttac
 181 agatgatttt gaatggaatt aataattaca agaatcccaa actcaccagg atgctcacat
 241 ttaagtttta catgccaag aaggccacag aactgaaaca gcttcagtgt ctagaagaag
 35 301 aactcaaacc tctggaggaa gtgctgaatt tagtcaaag caaaaacttt cacttaagac

361 ccagggaactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa
 421 cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga
 481 ttaccttttg tcaaagcacc atctcaacac taacttgata attaagtgtc tcccacttaa
 5 541 aacatatcag gccttctatt tatttattta aatatttaaa tttatattt attgttgaat
 601 gtatgggtgc tacctattgt aactattatt cttaatctta aaactataaa tatggatctt
 661 ttatgattct tttgtaagc cctaggggct ctaaaatggg ttaccttatt tatcccaaaa
 721 atatttatta ttatgttgaa tgttaaatat agtatctatg tagattgggt agtaaaacta
 781 ttaataaat ttgataaata taaaaaaaa aaacaaaaaa aaaaa (SEQ ID NO: 9)

10

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 47
- (2) 3' Untranslated Region - nts 519- 825

15 Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.4. Interleukin 6 ("IL-6")

20 GenBank Accession # NM_000600:

1 ttctgccctc gagcccaccg ggaacgaaag agaagctcta tctgcctcc aggagcccag
 61 ctatgaactc cttctccaca agcgccttcg gtccagtgc cttctccctg gggctgtctc
 121 tgggtgtgccc tgctgccttc cctgccccag tacccccagg agaagattcc aaagatgtag
 181 ccgccccaca cagacagcca ctacactctt cagaacgaat tgacaaacaa attcggtaca
 241 tctctgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa
 25 301 gcagcaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg
 361 gatgcttcca atctggattc aatgaggaga ctgacctgtt gaaaatcacc actggtcttt
 421 tggagtttga ggtataccta gactacctcc agaacagatt tgagagtagt gaggaacaag
 481 ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga
 541 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgtg acgaagctgc
 30 601 aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agctttaagg
 661 agttcctgca gtccagcctg agggctcttc ggcaaatgta gcatgggcac ctcagattgt
 721 tgttgtaaat gggcattcct tcttctggtc agaacctgt cactgggca cagaactat
 781 gttgttctct atggagaact aaaagtatga gcgttaggac actattttaa ttatttttaa
 841 ttattaata tttaaataatg tgaagctgag ttaatttatg taagtcatat ttatattttt
 35 901 aagaagtacc acttgaacaa tttatgtat tagttttgaa ataataatgg aaagtggcta

961 tgcagtttga atatcctttg tttcagagcc agatcatttc ttggaaagt taggcttacc
 1021 tcaaataaat ggctaactta tacatatttt taaagaaata ttatattgt atttatataa
 1081 tgtataaatg gttttatcac caataaatgg cattttaaaa aattc (SEQ ID NO: 11)

5

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 62
- (2) 3' Untranslated Region - nts 699 - 1125

10

Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.5. Vascular Endothelial Growth Factor ("VEGF")

15

GenBank Accession # AF022375:

20

25

30

35

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcgggcgt
 61 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc tttgggggt gaccgccgga
 121 gcgcggcggtg agccctcccc ctgggatcc cgcagctgac cagtgcgct gacggacaga
 181 cagacagaca cgcggccag cccagttac cacctctcc ccggccggcg gcggacagtg
 241 gacgcggcgg cgagccgcgg gcaggggccc gagcccggc ccggaggcgg ggtggagggg
 301 gtcggagctc gcggcgctgc actgaaactt ttcgtccaac ttctgggctg ttctgcttc
 361 ggaggagccg tggctccgcg gggggaagcc gagccgagcg gagccgcgag aagtgttagc
 421 tcgggcccggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag
 481 agggggccgc agtggcgact cggcgctcgg aagccgggct catggacggg tgaggcggcg
 541 gtgtgcgcag acagtgtcc agcgcgcgcg ctccccagcc ctggcccggc ctggggccgg
 601 gaggaagagt agctcgccga ggcgcccagg agagcgggccc gcccacagc ccgagccgga
 661 gagggacgcg agccgcgcgc cccggtcggg cctccgaaac catgaacttt ctgctgtct
 721 ggggtgattg gaggcttggc ttgtgtctct acctccacca tgccaagtgg tcccaggctg
 781 caccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct
 841 atcagcgag ctactgcat ccaatcgaga ccctggtgga catctccag gattaccctg
 901 atgagatcga gtacatcttc aagccatcct gtgtgcccct gatgcgatgc gggggctgct
 961 ccaatgacga gggcctggag tgtgtgccc ctgaggagtc caacatcacc atgcagatta
 1021 tgcggatcaa acctaccaa ggccagcaca taggagagat gagcttccta cagcacaaca
 1081 aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccttgct
 1141 cagagcggag aaagcatttg ttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa
 1201 acacacactc gcgttgcaag gcgaggcagc ttgagttaaa cgaacgtact tgcagatgtg

1261 acaagccgag gcggtgagcc gggcaggagg aaggagcctc cctcaggggt tcgggaacca
 1321 gatctctctc caggaaagac tgatacagaa cgategatac agaaaccacg ctgccgccac
 1381 cacaccatca ccatcgacag aacagtcctt aatccagaaa cctgaaatga aggaagagga
 5 1441 gactctgcgc agagcacttt gggtcaggag ggcgagactc cggcgggaagc attcccgggc.
 1501 gggtgaccca gcacggctcc tcttgaatt ggatcgcca tttattttt ctgtctgta
 1561 aatcaccgag cccggaagat tagagagttt tattctggg attcctgtag acacaccac
 1621 ccacatacat acatttatat atatatatat tatatatata taaaaataaa tatctctatt
 1681 ttatatatat aaaatatata tattcttttt ttaaattaac agtgctaag ttattggtgt
 10 1741 ctctactgga tgtatttgac tgctgtggac ttgagttggg aggggaatgt tccactcag
 1801 atcctgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtccact
 1861 tgggtggggcc agggctctct cccctgccc agaattgtga aggccagggc atggggggcaa
 1921 atatgaccca gttttgggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc
 1981 aggtcagacg gacagaaaga caaatcacag gtccgggat gaggacaccg gctctgacca
 15 2041 ggagtttggg gagcttcagg acattgctgt gctttgggga ticcctccac atgtcgcag
 2101 cgcatctgc cccaggggc actgcctgga agattcagga gcctgggcgg ccttcgcta
 2161 ctctcacctg ctctgagtt gccagaggg cactggcag atgtccggc gaagagaaga
 2221 gacacattgt tggaagaagc agccatgac agcgccctt cctgggactc gccctcatcc
 2281 tctctctgt ccccttctg gggtcagcc taaaaggacc tatgtctca caccattgaa
 20 2341 accactagtt ctgtccccc aggaacctg gttgtgtgtg tgtgagtggt tgaccttct
 2401 ccatccctg gtccttcct tccttccc aggcacagag agacaggga ggatccact
 2461 gccattgtg gaggcagaga aaagagaaag tgtttatat acggtactta ttaatatcc
 2521 cttttaatt agaaattaga acagttaatt taattaaaga gtagggtttt tttcagtat
 2581 tcttggttaa tatttaatt caactattta tgagatgtat ctttgctct ctctgctct
 25 2641 ctatttgta ccggttttg tatataaat tcatgttcc aatctctct tccctgatcg
 2701 gtgacagtca ctagctatc tgaacagat atttaatttt gctaacactc agctctgccc
 2761 tccccgatcc cctggctccc cagcacat tccttgaaa gagggtttca atatacatct
 2821 acatactata tatatattgg gcaacttgta ttgtgtgta tatatatata tatatgtta
 2881 tgtatatatg tgatctgaa aaaataaaca tcgtattct gtttttata tgttcaaacc
 2941 aaacaagaaa aatagagaa ttctacatac taaatctctc tccttttta attttaatat
 30 3001 ttgttatcat ttattattg gtgctactgt ttatccgtaa taattgtggg gaaaagatat
 3061 taacatcacg tctttgtct tagtgcagt ttccgagata ttccgtagta catatttatt
 3121 ttaaacaac gacaaagaaa tacagatata tcttaaaaa aaaaaa (SEQ ID NO: 12)

35 General Target Regions:

(1) 5' Untranslated Region - nts 1 - 701

(2) 3' Untranslated Region - nts 1275 - 3166

Initial Specific Target Motifs:

- 5 (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704
 5'CCGGGCUCAUGGACGGGUGAGGCGGCGGUGGCGCAGACAGU
 GCUCCAGCGCGCGCGCUCCCCAGCCUUGGCCCGGCCUCGGGGCCG
 GGAGGAAGAGUAGCUCGCCGAGGCGCCGAGGAGAGCGGGCCGC
 CCCACAGCCCGAGCCGGAGAGGGACGCGAGCCGCGCGCCCCGGU
 10 CGGGCCUCCGAAACCAUGAACUUUCUGCUGUCUUGGGUGCAUU
 GGAGCCUUGCCUUGCUGCUCUACCUCCACCAUG 3' (SEQ ID NO:
 13)
- (2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

15

6.6. Human Immunodeficiency Virus I ("HIV-1")

GenBank Accession # NC_001802:

1 ggtctctctg gtagaccag atctgagcct gggagctctc tggctaacta gggaaccac
 61 tgcttaagcc tcaataaagc ttgccttgag tgctcaagt agtgtgtgcc cgtctgtgt
 20 121 gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca
 181 gtggcgcccg aacagggacc tgaagcgaa agggaaacca gaggagctct ctgcacgcag
 241 gactcggcct gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc
 301 aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa
 361 gcgggggaga attagatcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat
 25 421 ataaattaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg
 481 gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc
 541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc
 601 atcaaaggat agagataaaa gacaccaagg aagcittaga caagatagag gaagagcaaa
 661 acaaaagtaa gaaaaagca cagcaagcag cagctgacac aggacacagc aatcaggta
 30 721 gccaaaatta ccctatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac
 781 ctagaacttt aatgcattg gtaaaagtag tagaagagaa ggctttcagc ccagaagtga
 841 taccatggtt tcagcatta tcagaaggag ccacccaca agatttaaacc accatgctaa
 901 acacagtggg gggacatcaa gcagccatgc aaatgtfaaa agagaccatc aatgaggaag
 961 ctgcagaatg ggatagagtg catccagtgc atgcagggcc tattgcacca ggccagatga
 35 1021 gagaaccaag gggaagtgc atagcaggaa ctactgtac ccttcaggaa caaataggat
 1081 ggatgacaaa taatccacct atccagtag gagaaattta taaaagatgg ataactctgg

1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
 1201 caaaggaacc ctttagagac tatgtagacc ggttctataa aactctaaga gccgagcaag
 1261 cttcacagga ggtaaaaaat tggatgacag aaacctgtt ggtccaaaat gcgaaccag
 5 1321 attgtaagac tattttaaaa gcatigggac cagcggctac actagaagaa atgatgacag
 1381 catgtcaggg agtaggagga cccggccata aggcaagagt ttggctgaa gcaatgagcc
 1441 aagtaacaaa ttcagctacc ataattgatc agagaggcaa ttttaggaac caaagaaaga
 1501 ttgtaagt ttcaattgt ggcaagaag ggcacacagc cagaaattgc agggccccta
 1561 ggaaaaagg ctgttgaaa tgtggaaagg aaggacacca aatgaaagt tgtactgaga
 1621 gacaggctaa ttttaggg aagatctggc ctctctaca gggaaggcca gggaatttc
 10 1681 ttcagagcag accagagcca acagccccac cagaagagag ctcaggtct ggggtagaga
 1741 caacaactcc cctcagaag caggagccga tagacaagga actgtatcct ttaactccc
 1801 tcaggtcact ctttgcaac gaccctcgt cacaataaag ataggggggc aactaaagga
 1861 agctctatta gatacaggag cagatgatac agtattagaa gaaatgagt tgcaggaag
 15 1921 atggaaacca aaaatgatag ggggaattgg aggtttatc aaagtaagac agtatgatca
 1981 gatactcata gaaatctgtg gacataaagc tataggta gtagtag gacctacac
 2041 tgtcaacata attggaagaa atctgttgac tcagattgt tgcacttaa atttcccat
 2101 tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaaggtta
 2161 acaatggcca ttgacagaag aaaaaataa agcattagta gaaatttga cagagatgga
 20 2221 aaaggaagg aaaatttcaa aaattggcc tgaaaatcca tacaatact cagtatttc
 2281 cataaagaaa aaagacagta ctaaatggag aaattagta gattcagag aacttaata
 2341 gagaactcaa gacttctggg aagtcaatt aggaatacca catccgcag ggttaaaaa
 2401 gaaaaaatca gtaacagtac tggatgtggg tgatgatat tttcagtic ctttagatga
 2461 agacttcagg aagtatactg catttaccat acctagtata acaatgaga caccaggat
 25 2521 tagatatcag tacaatgtc ttccacagg atggaaagga tcaccagca tattccaaag
 2581 tagcatgaca aaaatcttag agccttttag aaaacaaaat ccagacatag ttatctatca
 2641 atacatgat gatttgtat taggatctga ctagaata ggcagcata gaacaaaaat
 2701 agaggagctg agacaacatc tgttgagggt gggacttacc acaccagaca aaaaacatca
 2761 gaaagaacct ccattcctt ggatgggta tgaactccat cctgataat ggacagtaca
 30 2821 gcctatagt ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg
 2881 gaaattgaat tgggcaagtc agatttacc agggattaaa gtaaggcaat tatgtaaat
 2941 ccttagagga accaaagcac taacagaagt aataccata acagaagaag cagagctaga
 3001 actggcagaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgacctac
 3061 aaaagactta atagcagaaa tacagaagca ggggcaagg caatggacat atcaaattta
 3121 tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaattgagg gtgcccacac
 35 3181 taatgatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat

3241 atggggaaag actcctaaat ttaaactgcc catacaaaag gaaacatggg aaacatggtg
3301 gacagagtat tggcaagcca cctggattcc tgagtgggag ttgttaata cccctccctt
3361 agtgaaatta tggtaaccagt tagagaaaga acccatagta ggagcagaaa ccttctatgt
5 3421 agatggggca gctaacaggg agactaaatt aggaaaagca ggatatgtta ctaatagagg
3481 aagacaaaaa gttgtcacc taactgacac acaaatcag aagactgagt tacaagcaat
3541 ttatctagct ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc
3601 attaggaatc attcaagcac aaccagatca aagtgaatca gagtagtca atcaaataat
3661 agagcagtta ataaaaagg aaaaggctta tctggcatgg gtaccagcac acaaaggaat
10 3721 tggaggaaat gaacaagtag ataaattagt cagtgtgga atcaggaaag tactatttt
3781 agatggaata gataaggccc aagatgaaca tgagaaatat cacagtaatt ggagagcaat
3841 ggctagtgtat ttaacctgc cacctgtagt agcaaaagaa atagtagcca gctgtgataa
3901 atgtcagcta aaaggagaag ccatgcatgg acaagtagac tgtagtcag gaatatggca
3961 actagattgt acacatttag aaggaaaagt taccctggta gcagttcatg tagccagtgg
15 4021 atatatagaa gcagaagtta ttccagcaga aacagggcag gaaacagcat attttcttt
4081 aaaattagca ggaagatggc cagtaaaaac aatacatact gacaatggca gcaatttcac
4141 cgggtgctacg gttagggccg cctgttggtg ggccgggaatc aagcaggaat ttggaattcc
4201 ctacaatccc caaagtcaag gagtagtaga atctatgaat aaagaattaa agaaaattat
4261 aggacaggta agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat
20 4321 ccacaatttt aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga
4381 cataatagca acagacatac aaactaaaga attacaaaa caaattacaa aaaticaaaa
4441 ttttcgggtt tattacaggg acagcagaaa tccacttgg aaaggaccag caaagctcct
4501 ctggaaagggt gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaag
4561 aagaaaagca aagatcatta gggattatgg aaaacagatg gcagggtgatg attgtgtggc
25 4621 aagtagacag gatgaggatt agaacatgga aaagttagt aaaacaccat atgtatgtt
4681 cagggaagc taggggatgg ttttatagac atcactatga aagccctcat ccaagaataa
4741 gttcagaagt acacatccca ctaggggatg ctagattggt aataacaaca tattggggtc
4801 tgcatacagg agaaagagac tggcatttgg gtcagggagt ctccatagaa tggaggaaaa
4861 agagatatag cacacaagta gacctgaac tagcagacca actaattcat ctgtattact
30 4921 ttgactgttt ttcagactct gctataagaa aggccttatt aggacacata gttagcccta
4981 ggtgtgaata tcaagcagga cataacaagg taggatctct acaatacttg gactagcag
5041 cattaataac accaaaaaag ataaagccac ctttgcctag tgttacgaaa ctgacagagg
5101 atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggac
5161 actagagctt ttgaggagc ttaagaatga agctgttaga catttccta ggatttggct
35 5221 ccattggctta gggcaacata tctatgaac ttatggggat acttgggcag gattggaagc
5281 cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tctcgacata

5341 gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tcctagacta
 5401 gagccctgga agcatccagg aagtcagcct aaaactgctt gtaccaatig ctattgtaaa
 5461 aagtgttgct ttcattgcc aagttgttcc ataacaaaag ccttaggcat ctctatggc
 5 5521 aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct
 5581 ctatcaaagc agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca
 5641 ttagtagtag caataataat agcaatagtt gtgtgggtcca tagtaatcat agaatatagg
 5701 aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa
 5761 gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagatg
 10 5821 gggcaccatg ctcttggga tgttgatgat ctgtagtgt acagaaaaat tgtgggtcac
 5881 agtctattat ggggtacctg tgtggaagga agcaaccacc actctatttt gtgcatcaga
 5941 tgctaaagca tatgatacag aggtacataa tgtttgggcc acacatgcct gtgtaccac
 6001 agacccaac ccacaagaag tagtattggt aaatgtgaca gaaaatttta acatgtggaa
 6061 aaatgacatg gtagaacaga tgcattgagga tataatcagt ttatgggac aaagcctaaa
 15 6121 gccatgtgta aaattaaccc cactctgtgt tagtttaaag tgcactgatt tgaagaatga
 6181 tactaatacc aatagtagta gcgggagaat gataatggag aaaggagaga taaaaaactg
 6241 ctctttcaat atcagcaca gcataagagg taagggtcag aaagaatatg catttttta
 6301 taaacttgat ataataccaa tagataatga tactaccagc tataagtga caagttgtaa
 6361 cacctcagtc attacacagg cctgtccaaa ggtatccttt gagccaatic ccatacatta
 20 6421 ttgtgccccg gctggttttg cgattctaaa atgtaataat aagacgttca atggaacagg
 6481 accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac
 6541 tcaactgctg ttaaatggca gtctagcaga agaagaggta gtaattagat ctgtcaattt
 6601 cacggacaat gctaaaacca taatagtaca gctgaacaca tctgtagaaa ttaattgtac
 6661 aagacccaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt
 25 6721 tgttaacaata ggaaaaatag gaaatatgag acaagcacat tgtaacatta gtagagcaaa
 6781 atggaataac actttaaaac agatagctag caaattaaga gaacaatttg gaaataataa
 6841 aacaataatc ttaagcaat cctcaggagg ggaccagaa attgtaacgc acagttttaa
 6901 ttgtggaggg gaattttct actgtaatic aacacaactg ttaatatga cttggtttaa
 6961 tagtacttgg agtactgaag ggtcaataa cactgaagga agtgacaca tcacctccc
 30 7021 atgcagaata aaacaatta taaacatgtg gcagaaagta ggaaagcaa tgatgcccc
 7081 tccatcagt ggacaaatta gatgttcatc aaatattaca gggctgctat taacaagaga
 7141 tgggtgtaat agcaacaatg agtccgagat ctacagacct ggaggaggag atatgaggga
 7201 caattggaga agtgaattat ataataataa agtagtaaaa attgaacct taggagtagc
 7261 accaccaag gcaaagagaa gagtgggtgca gagagaaaaa agagcagtgg gaataggagc
 35 7321 ttgttcctt gggttcttgg gacgagcagg aagcactatg ggcgcagcct caatgacgt
 7381 gacggtagag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag

7441 ggctattgag gcgcaacagc atctgttgca actcacagtc tggggcatca agcagctcca
 7501 ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg ggatttgggg
 7561 ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgctagtt ggagtaataa
 5 7621 atctctggaa cagatttggg atcacacgac ctggatggag tgggacagag aaattaacaa
 7681 ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga
 7741 acaagaatta ttggaattag ataaatgggc aagtttggg aattggttta acataacaaa
 7801 ttggctgtgg tatataaat tattcataat gatagtagga ggcttggtag gtttaagaat
 7861 agtttttctt gtactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt
 10 7921 tcagaccac ctccaaccc cgagggggacc cgacaggccc gaaggaatag aagaagaagg
 7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg
 8041 ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactcttcat
 8101 tgtaacgagg attgtggaac ttctgggacg caggggggtg gaagccctca aatattggtg
 8161 gaatctccta cagtattgga gtcaggaact aaagaatagt gctgttagct tgctcaatgc
 15 8221 cacagccata gcagtagctg aggggacaga taggggtata gaagtagtac aaggagcttg
 8281 tagagctatt cgccacatac ctagaagaat aagacagggc ttggaaagga ttttctata
 8341 agatgggtgg caagtggta aaaagtagtg tgattggatg gcctactgta agggaaagaa
 8401 tgagacgagc tgagccagca gcagataggg tgggagcagc atctcgagac ctggaaaaac
 8461 atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttggtcc tggctagaag
 20 8521 cacaagagga ggaggaggtg ggttttcag tcacacctca ggtaccttta agaccaatga
 8581 cttacaaggc agctgtagat cttagccact ttttaaaga aaagggggga ctggaagggc
 8641 taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct
 8701 acttcctga ttagcagaac tacacaccag ggccaggggt cagatatcca ctgacctttg
 8761 gatggtgcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag
 25 8821 agaacaccag cttgttacac cctgtgagcc tgcattggat ggatgacctg gagagagaag
 8881 tgtagagtg gaggtttgac agccgcctag catticatca cgtggcccga gagctgcatc
 8941 cggagtactt caagaactgc tgacatcgag ctgtctacaa gggactttcc gctggggact
 9001 ttccaggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tctgcatat
 9061 aagcagctgc ttttgcctg tactgggtct ctctggtag accagatctg agcctgggag
 30 9121 ctctctggct aactagggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgtt
 9181 c (SEQ ID NO: 14)

Initial Specific Target Motifs:

- (1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1
 - 60
 35 "Minimal" TAR RNA element

5' GGCAGAUUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)

(2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element

5'

5

UUUUUUAGGGAAGAUCUGGCCUCCUACAAGGGAAGGCCAGG
GAAUUUUCUU 3' (SEQ ID NO: 16)

6.7. Hepatitis C Virus ("HCV" - Genotypes 1a & 1b)

GenBank Accession # NC_001433:

10 1 ttgggggcca cactccacca tagatcactc cctgtgagg aactactgtc ttcacgcaga
61 aagcgtctag ccatggcgtt agtatgagtg ttgtgcagcc tccaggaccc cccctcccgg
121 gagagccata gtggtctgcg gaaccggtga gtacaccgga attgccagga cgaccgggtc
181 ctttcttga tcaaccgct caatgcctgg agattgggc gtgccccgc gagactgcta
241 gccgagtagt gttgggtcgc gaaaggcctt gtgtactgc ctgatagggt gcttgcgagt
15 301 gccccgggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaagaaaaa
361 ccaaactaa caccaaccgc cgcccacagg acgttaagt cccgggcggt ggtcagatcg
421 ttggtggagt ttacctgttg ccgcgcaggg gccccaggtt ggggtgtcgc gcgactagga
481 agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatcccaag gctcgcggc
541 ccgagggtag gacctgggct cagcccggtt acccttgcc cctctatggc aacgagggta
20 601 tggggtgggc aggatggctc ctgtacccc gtggtctcg gcctagtgg ggccccacag
661 accccggcg taggtcgtt aatttggtta aggtcatcga taccctaca tgcggcttcg
721 ccgacctcat ggggtacatt ccgctgtcg gcgccccct agggggcgct gccagggccc
781 tggcacatgg tgcggggtt ctggaggacg gcgtgaacta tgcaacaggg aatctgccc
841 gttgctctt ctctatctt ctcttagctt tgctgtctt ttgacctc ccagctccg
25 901 cttacgaggt gcgcaacgtg tccgggatat accatgtcac gaacgactgc tcaactcaa
961 gtattgtgta tgaggcagcg gacatgatca tgcacacccc cgggtgcgtg ccctgcgtcc
1021 gggagagtaa ttctcccgt tgctgggtag cgctcactcc cagctcgcg gccaggaaca
1081 gcagcatccc caccacgaca atacgacgcc acgtcgattt gctcgttggg gcggctgctc
1141 tctgtccgc tatgtacgtt ggggatctct cgggatccgt tttctcgtc tccagctgt
30 1201 tcacctctc acctgccgg tatgagacgg tacaagattg caattgtca atctatccc
1261 gccacgtatc aggtcacgc atggcttggg atatgatgat gaactggtca cctacaacgg
1321 ccctagtgtg atcgcagcta ctccggatcc cacaagccgt cgtggacatg gtggcggggg
1381 cccactgggg tgcctagcg ggccttgctt actattccat ggtggggaac tgggctaagg
1441 tcttgattgt gatgtactc ttgctggcg ttgacgggca caccacgtg acagggggaa
35 1501 gggtagcctc cagcaccag agcctcgtt cctggctctc acaaggccca tctcagaaaa
1561 tcaactcgt gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg

1621 actccctcca aactgggttc attgctgcgc tgtctacgc acacaggttc aacgcgtccg
 1681 ggtgccccaga gcgcatggct agctgccgcc ccatcgatga gttcgctcag ggggtggggtc
 1741 ccatactca tgatatgcct gagagctcgg accagaggcc atattgctgg cactacgcgc
 5 1801 ctgaccgtg cgggatcgtg cctgcgtcgc aggtgtgtgg tccagtgtat tgcctcactc
 1861 cgagccctgt ttagtgggg acgaccgac gttcgggc tcctacgtat agctgggggg
 1921 agaatagac agacgtgctg ctacttagca acacgggcc gcctcaaggc aactggttg
 1981 ggtgcacgtg gatgaacagc actgggttca ccaagacgtg cggggggccct ccgtgcaaca
 2041 tcgggggggt cggcaacaac accttggtct gcccacgga ttgctccgg aagcaccgcg
 10 2101 aggccactta cacaagtgt ggctcggggc cctggtgac acccaggtgc atggtgact
 2161 acccatacag gctctggcac taccctgca ctgttaact taccgtctt aaggtcagga
 2221 tgtatgtggg gggcgtggag cacaggctca atgctcatg caattggact cgaggagagc
 2281 gctgtgactt ggaggacagg gataggtcag aactcagccc gctgctgctg tctacaacg
 2341 agtggcagat actgccctgt tcctcacca ccctaccggc cctgtccact ggcttgatcc
 15 2401 atcttcaccg gaacatcgtg gacgtgcaat acctgtacgg tatagggtcg gcagtgtct
 2461 cctttgcaat caaatgggag tatatcctgt tgccttct tctctggcg gacgcgcgcg
 2521 tctgtccctg cttgtggatg atgctgctga tagcccaggc tgaggccacc ttagagaacc
 2581 tgggtgtcct caatgcggcg tctgtggccg gagcgcatgg ccttctctcc ttctcgtgt
 2641 tctctgcgc cgctgttac atcaaaggca ggctggccc tggggcggca tatgtctct
 20 2701 atggcgatg gccgtgtc ctgctctgc tggccttacc accacgagct tatgccatgg
 2761 accgagagat ggctgcatc tgcggaggcg cggttttgt aggtctggtg ctctgacct
 2821 tgtcaccata ctataagggt ttctcgcta ggctcatatg gtggtacaâ tattttatca
 2881 ccagagccga ggcgcacttg caagtgtggg tccccctct caatgttcgg ggaggccgcg
 2941 atgccatcat cctccttaca tgcgcggtcc atccagagct aatctttgac atcaccaaac
 25 3001 tctgtctgc catactcggg ccgtcatgg tgcctcaggc tggcataact agagtccgt
 3061 actttgtacg cgctcagggg ctcatccgtg catgcatgtt agtgcggaag gtcgctggag
 3121 gccactatgt ccaaatggcc tcatgaagc tggccgcgt gacaggtagc tacgtatatg
 3181 accatcttac tccactcggg gattggggcc acgcgggcct acgagacctt gcggtggcag
 3241 tagagcccgt cgtctctct gacatggaga ctactcat cactggggg gcagacaccg
 30 3301 cggcgtgtgg ggacatcgc tcgggtctac cagtctccgc ccgaaggggg aaggagatac
 3361 ttctaggacc ggccgatagt ttggagagc aggggtggcg gctccttgc cctatcacgg
 3421 cctattccca acaaacgcgg gccctgctt gctgtatcat cactagcctc acagtcggg
 3481 acaagaacca ggtcgatggg gaggttcagg tgctctccac cgcaacgcaa tcttctgtg
 3541 cgacctgct caatggcgtg tgttgaccg tctaccatgg tgccggtcg aagaccctgg
 35 3601 ccggcccgaa ggtccaatc acccaatgt acaccaatgt agaccaggac ctgctcggct
 3661 ggccggcgcc ccccgggggc cgctccatga caccgtgcac ctgcggcagc tcggacctt

3721 acttgggtcac gaggcattgct gatgtcgttc cgggtgcgccg gcggggcgac agcaggggga
 3781 gcctgctttc cccagggccc atctcctacc tgaagggctc ctgggtgga cactgcttt
 3841 gcccttcggg gcacgttgta ggcatcttc cggctgctgt gtgcacccgg ggggtgcga
 5 3901 aggcgggtgga ctcatatccc gttgagtcta tggaaactac catgcggtct cggctctca
 3961 cagacaactc atccccctcg gccgtaccgc aaacattcca agtggcacat ttacacgctc
 4021 cactggcag cggcaagagc accaaagtgc cggctgcata tgcagcccaa ggttacaagg
 4081 tgctcgtcct aaaccgctcc gttcccgcca cattgggctt tggagcgtat atgtccaagg
 4141 cacatggcat cgagcctaac atcagaactg gggttaaggac catcaccacg ggcggcccca
 10 4201 tcacgtactc cacctattgc aagttccttg ccgacgggtg atgtccggg ggcgcctatg
 4261 acatcataat atgtgatgaa tgccactcaa ctgactcgac taccatcttg ggcacggca
 4321 cagtcctgga tcaggcagag acgggtggag cgcggctcgt cgtgctgcc accgccacgc
 4381 ctccgggatc gatcaccgtg ccacacccca acatcgagga agtggccctg tccaacactg
 4441 gagagattcc ctctatggc aaagccatcc ccattgaggc catcaagggg ggaaggcatc
 15 4501 tcattcttg ccattccaag aagaagtgtg acgagctcgc cgaaagctg acaggcctcg
 4561 gactcaatgc ttagcgtat taccggggtc tcatgtgtc cgtcataccg actagcggag
 4621 acgtcgttgt cgtggcaaca gacgtctaa tgacgggtt taccggcgac ttgactcag
 4681 tgatcgactg caacacatgt gtcaccaga cagtcgatt cagcttggat cccacctca
 4741 ccattgagac gacaacgctg cccaagacg cgggtgcgcg tgcgcagcgg cgaggtagga
 20 4801 ctggcagggg caggagtggc atctacaggt ttgtactcc aggagaacgg cctcaggca
 4861 tttcgactc ctgggtcctg tgtgagtgt atgacgcagg ctgcgttgg tatgagctca
 4921 cgcccgtga gacctcggtt aggttgcggg cttaactaaa tacaccaggg ttgccgtct
 4981 gccaggacca ctagagttc tgggagagcg tcttcacagg cctcaccac atagatgcc
 5041 acttcttgc ccagacaaa caggcaggag acaacctccc ctacctgta gcatacgaag
 25 5101 ccacagtgt cgccagggt caggctccac ctccatcgt ggaccaaatg tggagtgtc
 5161 tcatacggct aaagcccaca ctgcatgggc caacgccct gctgtacagg ctaggagccg
 5221 ttcaaatga ggtcactct acacaccca taacaaata catcatggca tgcattcgg
 5281 ctgacctgga ggtcgtcact agcacctggg tgctagtagg cggagtcctt gcggctctgg
 5341 ccgctactg cctgacgaca ggcagcgtg tcattgtggg caggatcatc ttgtccggga
 30 5401 ggccagctgt tattcccac aggggaagtc tctaccagga gttgatgag atggaagagt
 5461 gtgcttaca cctccctac atcgagcaag gaatgcagct cgccgagcaa ttcaacaga
 5521 aggcgtcgg attgtgcaa acagccacca agcaagcggg ggtgctgct cccgtgttg
 5581 agtccaagt gcgagccctt gaggtctct gggcgaaaca catgtggaac ttcatcagc
 5641 ggatacagta cttggcaggc ctatccactc tgcctggaaa ccccgcgata gcatcattga
 35 5701 tggcttttac agccttata accagcccgc tcaccacca aaataccctc ctgtttaaca
 5761 tcttgggggg atgggtggct gcccaactcg ctccccag cgtgcttcg gcttctgtg

5821 gcgccggcat tgccggtgcg gccgttgca gcataggtct cgggaaggta cttgtggaca
5881 ttctggcggg ctatggggcg ggggtggctg gcgcactcgt ggcctttaag gtcagagcg
5941 gcgagatgcc ctccactgag gatctggta atttactccc tgccatcctt tctctggcg
5 6001 ccctggttgt cggggtcgtg tgcgcagcaa tactgcgtcg gcacgtgggc ccgggagagg
6061 gggctgtgca gtggatgaac cggctgatag cgttcgttc gcggggtaac cacgtctccc
6121 ccacgcacta tgtgcccag agcgacgccg cggcgcgtgt tactcagatc ctctccagcc
6181 ttaccatcac tcagttgctg aagaggcttc atcagtggat taatgaggac tgctccagc
6241 cttgttcggg ctctggcta aaggatgtt gggactggat atgcacggtg ttgagtact
10 6301 tcaagacttg gctccagtc aagctcctgc cgcggttacc gggactccct ttctgtcat
6361 gccaacgcgg gtacaaggga gtctggcggg gggatggcat catgcaaacc acctgccat
6421 gtggagcaca gatcaccgga catgtcaaaa atggctccat gaggattgtt gggccaaaa
6481 cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc atacaccag gggccctgca
6541 cgccctcccc agcgccgaac tattccaggg cgctgtggcg ggtggctgct gaggagtacg
15 6601 tggaggttac gcgggtgggg gatttcact acgtacggg catgaccact gacaacgtga
6661 aatgcccatg ccaggttcca gcccctgaat ttctacgga ggtggatgga gtacggtgc
6721 acaggatgac tccagtgtgc aaacctctcc tacgagagga ggtcgtattc caggctgggc
6781 tcaaccagta cctggtcggg tcacagctcc catgtgagcc cgaaccggat gtggcagtg
6841 tcacttccat gctcaccgac ccctctcata ttacagcaga gacggccaag cgtaggctgg
20 6901 ccagggggtc tccccctcc ttggccagct cttcagctag ccagttgtct gcgccttctt
6961 tgaaggcgac atgtactacc catcatgact ccccgagcgc tgacctatc gaggccaacc
7021 tctgtggcg gcaggagatg ggcgggaaca tcaccgtgt ggagtcagaa aataaggtgg
7081 taatcctgga ctcttctgat ccgattcggg cgggtggagga tgagaggga atatccgtc
7141 cggcggagat cctgcgaaaa cccaggaagt tccccccagc gttgccata tgggcacgcc
25 7201 cggattacaa ccctccactg ctagatcct ggaaggaccc ggactacgtc ccccggtg
7261 tacacgggtg cctttgcca tctaccaagg ccccccaat accacctcca cggaggaaga
7321 ggacgggtgt cctgacagag tccaccgtgt cttctgcctt ggcggagctc gctactaaga
7381 ctttggcag ctccgggtcg tcggccgtg acagcggcac ggcgactggc cctccgac
7441 aggcctccga cgacggcgac aaaggatccg acgttagtc gtactcctcc atgcccccc
30 7501 tcgagggaga gccaggggac cccgacctca gcgacgggtc ttggtctacc gtgagcggg
7561 aagctggtga ggacgtcgtc tgctgtcaa tgcctatac atggacaggt gccttgatca
7621 cgccatgcgc tgcggaggag agcaagtgc ccatcaatcc gttgagcaac tcttctgctc
7681 gtcaccacag tatggtctac tcacaacat ctgcagcgc aagtctcgg cagaagaagg
7741 tcaccttga cagactgcaa gtctggacg accactaccg ggacgtgctc aaggagatga
35 7801 aggcgaaggc gtccacagtt aaggetaggc ttctatctat agaggaggcc tgcaactga
7861 cgccccaca ttcggccaaa tcaaatttg gctacggggc gaaggacgtc cggagcctat

7921 ccagcagggc cgtaaccac atccgctccg tgtgggagga cttgctggaa gacactgaaa
 7981 caccaattga taccaccatc atggcaaaaa atgaggtttt ctgctgcaa ccagagaaag
 8041 gagggcgcaa gccagctcgc cttatcgtat tcccagacct ggggggtacgt gtatgcgaga
 5 8101 agatggccct ttacgacgtg gtctccacc ttctcaggc cgtgatgggc ccctcatag
 8161 gattccagta ctctcctggg cagcgggtcg agttcctggt gaatacctgg aaatcaaga
 8221 aatgccctat gggtcttca tatgacacc gctgcttga ctcaacggtc actgagaatg
 8281 acatccgtac tgaggaatca attaccaat gttgtgactt ggccccgaa gccaggcagg
 8341 ccataaggtc gctcacagag cggctttatg tgggggtcc cctgactaat tcgaaggggc
 10 8401 agaactgagg ttatgcggg tgccgcgcaa gtggcgtgct gacgactagc tgcggcaaca
 8461 cctcacatg ttactgaag gccactgagg cctgtcagc tgcaaagctc caggactgca
 8521 cgatgctcgt gaacggagac gacctgtcg ttatctgtga gagtgcggga acccaggagg
 8581 atgcggcggc cctacgagcc ttacggagg ctatgactag gtattccgcc cccccgggg
 8641 accgccccca accagaatac gacttgagc tgataacgtc atgctctcc aatgtgtcgg
 15 8701 tcgcgcacga tgcacccgc aaaaggggtg actacctac ccgtgacccc accaccccc
 8761 tcgcacgggc tgcgtgggag acagttagac aactccagt caactcctgg ctaggcaata
 8821 tcactatgta tgcgccacc ctatgggca ggtatgtct gatgactcat ttcttcta
 8881 tcctctagc tcaggagcaa cttgaaaag ccctggattg tcagatctac ggggcctgtt
 8941 actccattga gccactgac ctacctaga tcattgaacg actccatggt cttagcgtat
 20 9001 ttctactcca cagttactct ccagggtgaga tcaatagggt ggcttcatgc ctcaggaaac
 9061 ttgggggtacc gcctttgca gtctggagac atcgggccag aagtgtccgc gctaagctac
 9121 tgtccagggt ggggagggct gccactgag gcaagtacct cttaactgg gcagtaaaga
 9181 ccaagcttaa actcactcca atccggctg cgtccagct agactgtcc ggctggttcg
 9241 ttgctggta caacggggga gacatatac acagcctgc tegtccga cccgtgtg
 25 9301 tcattgttg cctactcta cttctgtg ggtaggcat ctactgtc cccaaccgt
 9361 gaacggggag ctaaccactc caggccaata ggccattccc tttttttt ttc (SEQ ID NO: 17)

General Target Region:

5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

30 5'UUGGGGGCGACACUCCACCAUAGAUCACUCCCCUGUGAGGAACUACUGUCU
 UCACGCAGAAAGCGUCUAGCCAUGGCGUUAGUAUGAGUGUUGUGCAGCCUC
 CAGGACCCCCCUCCGGGAGAGCCAUGUGGUCUGCGGAACCGGUGAGUAC
 ACCGGAUUGCCAGGACGACCGGGUCCUUUCUUGGAUCAACCCGCUCAAUGC
 CUGGAGAUUUGGGCGUGCCCCGCGAGACUGCUAGCCGAGUAGUGUUGGGU
 35 CGCGAAAGGCCUUGUGGUACUGCCUGAUAGGGUGCUUGCGAGUGCCCCGGG
 AGGUCUCGUAGACCGUGCAU3' (SEQ ID NO: 18)

Initial Specific Target Motifs:

- (1) Subdomain IIIc within HCV IRES - nts 213 - 226
 5'AUUUGGGCGUGCCC3' (SEQ ID NO: 19)
- (2) Subdomain IIIId within HCV IRES - nts 241-267
 5'GCCGAGUAGUGUUGGGUCGCGAAAGGC3' (SEQ ID NO: 20)

6.8. Ribonuclease P RNA ("RNaseP")

10 GenBank Accession #s

X15624 Homo sapiens RNaseP H1 RNA:

1 atgggaggag ggaagctcat cagtggggcc acgagctgag tgcgtcctgt cactccactc
 61 ccatgtccct tgggaaggct tgagactagg gccagaggcg gccctaacag ggctctccct
 121 gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggctcag actgggcagg
 15 181 agatgccgtg gaccccgccc ttcggggagg ggcccggcgg atgcctcctt tccgggagct
 241 tggaacagac tcacggccag cgaagtgagt tcaatggctg aggtgaggta ccccgaggg
 301 gacctcataa cccaattcag accactctcc tccgccatt (SEQ ID NO: 21)

U64885 Staphylococcus aureus RNaseP (rrnB) RNA:

20 1 gaggaaagtc cgggctcaca cagtctgaga tgattgtagt gttcgtgctt gatgaaacaa
 61 taaatcaagg cattaatttg acggcaatga aatatcctaa gtcttcgat atggatagag
 121 taatttgaaa gtgccacagt gacgtagctt ttatagaaat ataaaagggtg gaacgcggta
 181 aaccctcga gtgagcaatc caaatttggg aggagcactt gttaacgga attcaacgta
 241 taaacgagac acacttcgag aatgaagtg gtgtagacag atggttatca cctgagtacc
 25 301 agtgtgacta gtgcacgtga tgagtacgat ggaacagaac gcggcttat (SEQ ID NO: 22)

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P (rnpB) gene:

30 1 gaagctgacc agacagtcgc cgcttcgtcg tcgtcctctt cgggggagac gggcggaggg
 61 gaggaaagtc cgggctccat agggcagggt gccaggtaac gcctgggggg gaaacccacg
 121 accagtcaa cagagagcaa accgccgatg gcccgcgcaa gcgggatcag gtaagggtga
 181 aagggtgcgg taagagcgca ccgcgcggct ggtaacagtc cgtggcacgg taaactccac
 241 ccggagcaag gccaaatagg gggtcataag gtacggcccg tactgaaccc gggtaggctg
 301 cttagaccag tgagcgattg ctggcctaga tgaatgactg tccacgacag aacccggctt
 35 361 atcggtcagt ttacct (SEQ ID NO: 23)

Z70692 *Mycobacterium tuberculosis* RNaseP (rnpB) RNA:

1 ccaccgggta cgatcttgcc gaccatggcc ccacaatagg gccggggaga cccggcggtca
 61 gtggtgggcg gcacggtcag taacgtctgc gcaacacggg gttgactgac gggcaatata
 121 ggctccatag cgtcggccgc ggatacagta aaggagcatt ctgtgacgga aaagacgcc
 181 gacgacgtct taaaacttg caggacgag aaggatgaat atgtcgacgt ccggttctgt
 241 gacctgctg gcatcatgca gcaattcacg attccggctt cggccttga caagagcgtg
 301 ttgacgacg gcttggcctt tgacggctcg tcgattcgcg ggtccagtc gatccacgaa
 361 tccgacatgt tgctcttcc cgatcccgag acggcgcgca tcgacccgtt ccgcgcggcc
 421 aagacgctga atatcaactt cttgtgcac gaccggttca ccctggagcc gtactcccgc
 481 gaccgcgca acatgcccc caaggccgag aactacctga tcagcactgg catcgccgac
 541 accgcatact tcggcgccga ggccgagttc tacatttcg attcggtag cttcgactcg
 601 cgcgccaacg gctccttcta cgaggtggac gccatctcgg ggtggtggaa caccggcgcg
 661 gcgaccgagg ccgacggcag tccaaccgg ggctacaagg tccgccacaa gggcgggtat
 721 ttccagtgg ccccaacga ccaatcgtc gacctgcgcg acaagatgct gaccaacctg
 781 atcaactccg gcttcatcct ggagaagggc caccacgagg tgggcagcgg cggacaggcc
 841 gagatcaact accagttaa ttcgtgctg cagccgcgcg acgacatgca gttgtacaag
 901 tacatcatca agaacaccgc ctggcagaac ggcaaacgg tcacgttcat gcccaagccg
 961 ctgttcggcg acaacgggtc cggcatgcac tgtcatcagt cgctgtggaa ggacggggcg
 1021 ccgctgatgt acgacgagac gggttatgcc ggtctgtcgg acacggcccg tcattacatc
 1081 ggcgccctgt tacaccacgc gccgtcgtg ctggccttca ccaaccgac ggtgaactcc
 1141 tacaagcggc tgggtcccgg ttacgaggcc ccgatcaacc tggctatag ccagcgcaac
 1201 cggtcggcat gcgtgcgcat cccgatcacc ggcagcaacc cgaaggccaa gcggctggag
 1261 ttccgaagcc ccgactcgtc gggcaaccgg tatctggcgt tctcgccat gctgatggca
 1321 ggcctggacg gtatcaagaa caagatcgag ccgcaggcgc ccgtcgacaa ggatctctac
 1381 gagctgccgc cggaagaggc cgcgagtatc ccgcagactc cgaccagct gtcagatgtg
 1441 atcgaccgtc tcgaggccga ccacgaatac ctaccgaag gaggggtgtt cacaacgac
 1501 ctgatcgaga cgtggatcag ttcaagcgc gaaaacgaga tcgagccggt caacatccgg
 1561 ccgcatccct acgaattcgc gctgtactac gacgtttaag gactcttcgc agtccgggtg
 1621 tagaggggagc ggcgtgtcgt tgccagggcg ggcgtcgagg ttttcgatg ggtgacggtg
 1681 gccggcaacg gcgcgccgac caccgtcgcg aagagcccgt ttaagaactg tcaaggacgt
 1741 ttacccggg tgccacaacc cgcttggcaa tcattcccc accgccgagc gggttgtctt
 1801 tcacatgcgc cgaaactcaa gccacgtcgt cggccaggcg tgctcgtcgc gccggttcag
 1861 gttaatgtc ggggattcgt cgtcggggcg ggcgtccacg ctgaccaacg gggcagtcga
 1921 ctcccgaaca ctttgcgcac taccgcctt gcccgccgcg tcaccgtag gtagttgtcc
 1981 aggaattccc caccgtcgtc gtttcgccag ccggccgcga ccgcgaccgc attgagctgg

2041 cgcccggtg cggcagctg gtcggtggg ttgccgcga ccaacaccag cgcgttgccg
 2101 gcccggtg cggcagcca ggcctgacgg agcagctcca cgtcggctgc gggaaccaga
 2161 tcggcgccg cgtgacatc cagggttgc agcgtcagg tttgtgcag ggcgggaacc
 5 2221 tggcgcat gctgtagt cagcaactgc acggtccatt cgtgctggc cagtccgcc
 2281 cggcccagtt tgggtgtgt gtggggctg gcaccgcgc gcaaccgctc ggactcgata
 2341 cgggccttga tggcggaat ctgcgcacc ggtcagcgg acacaccgtc ggcgggatac
 2401 cgcgtttgt caccatccg taggaatgc tgaccaact cggcatgcc ggcaaccgcg
 2461 tgtgcgta gcaggcctg gatcccat ggtgtgccc actgctgta gtatcgggc
 10 2521 taggaccca ggtgaggac cagcgaccg ttgcggcct cgggtgcaa attggcgtc
 2581 agtccagcg gcgtagcag gtcgggtgc ccagcagcg ccgaaccg ctggcgatc
 2641 gatgtagc atttaccgc cgtgcatc tcgacccgg tggccggtc acagacgaac
 2701 atcagtcgg catccgacc gtagccaa tcggcaccac ccagccgacc catccgatg
 2761 acccgcatg ccgccggggc gcgtagctc tcgggaaggc tggccggat catgacgtc
 15 2821 agcggcgct gcagaccgc caccacacc gacgtcaac ccggcacac ctgggtgac
 2881 tcgagcagg cgagcaggc cgccgaacc atcgggcca gctctgac acgcagcgtg
 2941 cgcgcggcg cgtggccc ctcgggtc gggtagcgg tcgagaggc gatcagcgc
 3001 cgagccagg cggcggtc ggtctgagc agttcgggc ccgagggcc gtcctgtac
 3061 tgctggatg ccgcgggc gcgcatcaac agatccggc ctagccga ggtaccaag
 20 3121 acatgcatg gccgttggc caccgggg ttgtccgca cgtggccag gtaccagt
 3181 tcggtggca gcgctcact gagcccggg taggccagc gtcgcccgc gggatcggg
 3241 gcatacgaca tcagtcag cagctgggc agcagaccg actgcaccg tccgcccgg
 3301 ccgtttgat tgaccaacgc cagatgtgt tcaacggg tctcggtcc ctctagccc
 3361 agcgggcca gccggcgccc cgggctcc aacgtatgc cgtggcgat ctcaaccg
 25 3421 gtcggccga tcgattccag cagcggtga tagaagatt tgggtgtaa ctgcacac
 3481 cgcagttct gttcttgag ttctcccgc agcaccgg ccgcatgt tggccatc
 3541 ggccggatg gggccgcgc gccagccag cgcactgc cctcgttc gggatcgga
 3601 agcaggtgg tgcgttag ccgtgcaac tgagtcggt gtcgagcag cctgaggaa
 3661 tcatacgac cgtcatgt cgccgctc tcagcccga ttagccgc ttgcccac
 30 3721 gccgcaatg cgtccaccg ggacgccac cgtacgact cgtcgctac ggcatgaac
 3781 agtgcagta gctgtacgc gaactccag tcgcaatc ccgctgcc gatttgagc
 3841 tcgcgccgc ggacatggc gggcaccag tctccacc gcccgccat ggcctgcac
 3901 tcgaccaca agtcttcgc ctgcaggct cgccacca tcggcatca ggcggtcagg
 3961 taacgctgc caagtccgc gtcgccaac actggccgtg cttcagca cgctgaaac
 35 4021 tccaggtct tggccagcg ctggtagtag gcgtagtc actcgagct acggaccagc
 4081 tcccgttc gccctccg agcagggcg gcgtccact cgaagggc cgccgaggc

4141 acccgcatca tctcgctggc cacgcgcgcg ttgcgcgggt cggagcgctc ggcaacgaat
4201 atgacatcga cgtcgctgac gtagttcagt tcgcgcgcac cgcactigcc catcgcgatg
4261 accgccaggc ggggtggcgg gtgctcgccg cacacgctcg cctcggccac gcgcagcgcc
5 4321 gccgccagag cggcgtccgc ggcgtccgcc aggcgtgcgg ccaccacggt gaattggcagc
4381 accggttcgt cctcgaccgt cgcggccagg tcgagagcgg ccagcattag cacgtagtcg
4441 cgggtactggg ttgcaatcg gtgcacgagc gagcccggca taccctccga ttctcgacg
4501 cactcgacga acgaccgctg cagctggtca tgggacggca gtgtgacctt gccccgcagc
4561 aatttcagg actgcgatg ggcgaccagg tgatgccca acgccagcga cgagcccagc
10 4621 accgagaaca gccgcccgcg cagactgcgt tcgcgcagca gagccgcgtt gagctcgtcc
4681 catccggtgt ctggattctc cgacagccgg atcaaggcgc gcagcgcggc atcggcgtcc
4741 ggagcgcgtg acagcgacca cagcaggtcg acgtgcgctt gatcctcgtg ccgatccac
4801 cccagctgag ccagacgctc accagcaggg gggtaacta atccgagccg gccaacgctg
4861 ggcaacttcg gccgctgcgt ggcgagttg gtcacgacca cgacggtagc gcaaagcgcg
15 4921 tcggcgctgg atcaaccggt agatctgggc tacagcgaca ggtaggtgcg cagctcgtat
4981 ggcgtgacgt ggctgcggta gttgcccac tccgtgcgtt tgttgcgaa gaaaaagtca
5041 aaaacgtgct ccccaaggc ctcgcgcagc agttcggagg cctccatggc gcgcagcgca
5101 ctatccaaac tggacggcaa ttctcggtac cccatcgctc ggcgttcctc ggggtgtgagg
5161 tcccatactg tgtctcggc ctgcggggcc agcacgtaac ccttctctac accccgcaat
20 5221 cccgcggcca gcagcacggc gaattgtcaga tagggattgc acgccgaatc agggctgcgt
5281 acttcgaccc gccgcgacga ggtcttctgc ggcgtgtaca tcggcacccg cactagggcg
5341 gatcgggtgg cggccccca cgacggggcc gtggcgctt cgcgccttg caccagccg
5401 ttgtaagagt tgaccactg atttgtacc gcgtgatct cgcaagcgtg ctccaggatc
5461 ccggcgatga acgatttacc cacttcgac agctgcagcg gatcatcagc gctgtggaac
25 5521 gcgttgacat caccctcga caggctcatg tgggtgtgca tcgccagcc cgggtgtggtg
5581 ccgaatggct tgggcatgaa cgacggccg ggcgcctctt ccagcgcgac ttcttgatg
5641 acgtagcgga aggtcatcac gttgtcagc atcgacagag cgtcggcaaa ccgcaggtcg
5701 atctctgct ggcgggtgc gccttcgtga tggctgaact ccaccgagat gccatgaat
5761 tccagggcat cgatcgctg gcggcgaaag ttcaaggcgg agtcgtgcac cgcttggtcg
30 5821 aaatagccgg cgttgcgac cgggacgggc accgaccgt cctcgggtcc gggcttgagc
5881 aggaagaact cgatttcggg atgcacgtag caggagaagc cgagttcgcc ggccttcgct
5941 agtcgccgc gcaacacgtg ccgcgggtcc gccacgacg gcgagccgtc cggcatggtg
6001 atgtcgcaaa acatccgcgc tgagtgtggtg tggccggaac tgggtggcca gggcagcacc
6061 tggaaagtgc acgggtccg gtgcgccacc gtatcggatt ccgagaccg cgaagccc
35 6121 tcgatcgagg atccgtcga gccgatgcct tctcgaagg cgcctcgag ttgggtggg
6181 gcgatggcga ccgacttgag gaaaccgagc acgtctgtga accacagccg gacgaagcgg

6241 atgtcgcgtt cttccagggt acgaagaacg aattccttct gtcggtccat acctcgaaca
6301 gtatgcactg tctgttaaaa ccgtgttacc gatgcccggc cagaagcgtt gcggggcggc
6361 ccgcaagggg agtgcgcggt gagttcaggg cgcgcaccgc agactcgtcg gcggcaaggt
5 6421 cccgtcgaga aaatagtga tcaccgcaga gtccacacac tggttgccat cgaacaccgc
6481 agtgtgttgg gtgccgtcga aggtgatcag cgggtcgccc agctggcggg ccaggtctac
6541 cccggactga tacggagtgg ccgggtcgtg ggtggtggac accacgacga ccttgccagc
6601 cccggccggc gccgcggggt gcggcgtcga cgttgccggc accggccaca gcgcgcacag
6661 atcgcggggg gcggatccgg tgaactgccg gtagctaagg aacggggcga cctgacggt
10 6721 ccgttggtcg gcggccaccc aggccgttgg atcgcccggt gtggcgcat cgacgcaccg
6781 gaccgcgttg aacgcgtcct ggtcgttct gtagtcccc tctgcatccc ggccgtcata
6841 gtcgtcggca agcaccagca agtcgccggc gtcgtgccg cgctgcagcc ccagcagacc
6901 actggtcagg tacttccagc gctgagggt gtacagcgcg ttgatgtgc ccgtcgtcg
6961 gtcggcgtag ctcaggccac gtggatccga cgtcttacc ggcttctga ccagcgggtc
15 7021 aaccagggcg tggtagcggt tgaccactg ggccgagtcg gtgccagag ggcaggccgg
7081 cgagcggcg cagtcggcgg cgtagtcatt gaaagcggc tgaatccc ccaattgggt
7141 gatgcttcc tcgattggc taacggctgg atcgatagcg ccgtcgagga ccatgcccc
7201 cacatagta ccgaaccgtt ccaggtgaag ggtgccaac tcggtgccgt agctgtatcc
7261 gaggtagtig atctgatcgt cacctaacgc ttggcgaacc atgtccatgt cccgtgcgac
20 7321 ggacgcgga ccgatatgg ccaagaagct gaagcccatc cgttaacac agtctgggc
7381 caactgccgg tagacctgt cgacgtgggt gacaccggc ggactgtagt cggccatcgg
7441 atcgcgccgg tacgcgtcga actcggcgtc ggtgcgacac cgcaacgcag ggtcgagtg
7501 gccgaccct ctcgggtcga agcccaccag gtcgaagtgg cggagaatgt cgggtgcggc
7561 gatcgcgggt gccatagcgg cgacctgtc gaccgccgac gccccgggtc cccaggtat
25 7621 gaccagcagt gtcctgaatc gctgtcccgt cgcggggacg cggatcaccg ccaacttcg
7681 ttgtgtcca ccgggttgg cgtagtcgac ggggacggac accgtcgcgc agcgtgcagt
7741 gcgaatttcg ctggtgtcgg cgatgaatc gcggcagctg ttcaactct gttgcggcgc
7801 cagcaccggc gcaccggggg ttggccggc gccgggttct tcagtcgcgc cggccaacgg
7861 gggcgtgct aggggcagtc cgccgagcag caaccgaag gacagcagc cggagctcaa
30 7921 cggctgcgg gccacatgg ccgccatcgt ctcaccggcg aatactgtg acggcgcgaa
7981 atgatcacac cttcgttct tcgccccgt agcacttggc gccgtgggc ggcgtggtgc
8041 cgccgattaa atacgccgtc acgtactgt caatgcagct gtcgccctgg aataccaccg
8101 tgtgtgggt tccgtcgaag gtcagcaac aaccgcgaag ctggttcgcc aggtcgaccc
8161 cggccttcta cggcgtcgc ggtcatggg tgggtgatac caccaccgtc ggcactaggc
35 8221 cgggcgccga gacggcatgg ggctgactg tgggtggcac cggccagaac gcgcaggtgc
8281 ccagcggcgc ataccggtg aacttcccgt agctcatgaa cggtcgcatc tcccgggcgc

8341 ggcggtcttc gtcgatgacc ttgtcgcgat cggtaaccgg gggctgatcg acgcaattga
8401 tcgccaccgg cgcgtcaccg gaattgtgt agcggccgtg cgagtcccga cgcattgaca
8461 tgtcggccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgtcgg
5 8521 tcaagtgtt ccacagattc ggtgagtaca gcgccataat ggtgcccacg atggcgtcgc
8581 tataactcag cccgcgcgga tccttcgtgc gcgccggcct gctgatcctc gggttgtccg
8641 ggtcgaccaa cggatcgacc aggctgtggt agacctcgac ggctttggcc gggtcggcgc
8701 ccagcgggca gcccgcgttc ttggcgcagt cggcggcata gttgtgaac gcgtcctgga
8761 agcccttggc ctggcgcagc tccgcctcga tgggatcggc attggggtcg acggcacctg
10 8821 cgagaatcat tccccgacc cgtcgcgga attcctcggc atacgggag ccgatccggg
8881 tgccgtacga gtagccagg taggtcagct tgtcgtcgcc caacgccgcg cgaatggcat
8941 ccaggtcctt ggcgacgttg accgtcccga catgggccag aaagticttg cccatcttgt
9001 ccacacagcg accgacgaat tgcttggtct cgttctcgat gtgcgccaca cctcccggc
9061 ttagtcaac ctgcggctcg gcccgcagcc ggtcgtgtgc ggcatcgag ttgcaccaga
15 9121 tcgccggccg ggacgacgc accccgcggg ggtcgaacc aaccaggtcg aaccttcgt
9181 gcaccgcgtt cggcaatgtc tggaagacgc ccaaggcggc ctcgataccg gattcgccgg
9241 gtccaccggg atttatgacc agcgaaccga tctgtctcc cgtcgccga aagcgaatca
9301 gcgccagcgc cgccacgtca ccatcggggc ggtcgtatg gaccgttaca gcgagcttgc
9361 cgcataacgc gccgcggggg atctttactt gcgggttga cgaccggcac ggtgtccact
20 9421 ccaccggctg gccagcttc ggctccgcca tacgagcgcg tccccgacc acgcggatgc
9481 agcccacaag aaccaacgc accgcgcgga gcgcggccca gatcaacagc atgcgcgca
9541 tctgtcgcg gcgagacagc ctcattccca caatgctgcc agagcagacc cgagatcctg
9601 gccagcggcc accgtcggcc gactaaccgg ccgtgccag cagtctgcc atgcggatg
9661 gcgaactcgt cggccatccc ccatagtcg ggtaacagat ccgggcaaga caaccgacccg
25 9721 tcgaccggat ccggcacggg cgcgtcggcc tcggcgggtc acaactgca catcaggttg
9781 gcgttgccac cccgtccag ccggcatggt gcaccttggc catcgccga gggcgatccc
9841 cgatgccgtc cacccttcg acgaacccat ctcccacggc ggtcgccggc agcgacgca
9901 tgtggccgca gatctccgag agttcggccc gcccgccgg cgacggcaac ccgatgccgt
9961 gcaagtgcg atcgatgtga ggttcaaggt tcagcgact gctggcaagc ttttccgaa
30 10021 acccgggcct cgccttgatc tggagtaca acgcgtcacg cagccggtca aaggcgtaac
10081 ccatgctcga gcaaacatgc atgggctgag tggacgttc cagacacagc aactggcgtc
10141 caggccactg agccgtgca tgcgcatgg tatgccgatg ggggccccgg gcgcgtctga
10201 ggggaagaag tggcagactg tcagggtccg acgaaccgg ggaccctaac gggccacgag
10261 gatcgacccg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata
35 10321 ccccgaggag ctccggggcg cggaccaaga tccgcacca ccacctacag agatggaagg
10381 ccgacggcca caagtgggcc atgctgacgg cctacgacta ttcgacggcc cggatcttcg

10441 acgaggccgg catcccgggtg ctgctggctg gtgattcggc ggccaacgtc gtgtacggct
10501 acgacaccac cgtgccgata tccatcgacg agctgatccc gctgggtccgt ggcgtgggtg
10561 ggggtgcccc gcacgcactg gtcgtgccg acctgccgtt cggcagctac gaggcggggc
5 10621 ccaccgccgc gttggccgcc gccaccgggt tcctcaagga cggcggcgca catgcggta
10681 agctcgaggg cggtagcgg gttggccgagc aaatcgctg tctgaccgcg gcgggcatcc
10741 cggtagtggtg acacatcggc ttcacccgc aaagcgtcaa cacctgggc ggttcggg
10801 tgcagggccg cggcgacgcc gccgaacaaa ccatcgcca cgcgatgcc gtcgccgaag
10861 ccggagcgtt tgccgtcgtg atggagatgg tgccgccga gttggccacc cagatcacc
10 10921 gcaagcttac cattccgacg gtcgggatcg gcgctgggcc caactcgac ggccaggtcc
10981 tggtaggca ggacatggc ggggtcagc gcgccaagac cggccgttc gtaaacggt
11041 atgccgatgt cggtagtgaa ctacccgtg ctgcaatga atagcccaa gagtgggc
11101 gcgggggtatt ccccgctgac gaacacagt tctgaccaag ccgaatcagc ccgatgcgc
11161 ggcatcgcg tggcgccctg gatgccgtc acgccggatt gccggcgcg acgcccagc
15 11221 gggaccatc ggcgtcgtg tgcgggtg agccggggg gagccagac attgatgtg
11281 cccaacacca tccgccagc ccaattgat gtggcactc atgatgct atccccgac
11341 aaccaccacc gcggcgacgc atcatgaccg gaggcgaaga tgccagtaga ggcccccaga
11401 ccagcgccg atctggaggt cgagcgcaag ttcgacgtga tcgagtcgac ggtgtcggc
11461 tcttcgagg gcatcgccg ggtggttcg gtcgagcagt cggcgacca gcagtcgac
20 11521 gcggtgtact tcgacacacc gtcgcacgac ctggcgcgca accagatcac cttggcgcg
11581 cgcaccggcg gcgccgacgc cggctggcat ctgaagctg cggccggacc cgacaagcg
11641 accgagatg gagcaccgt gtccgatca ggcgacgtg tgccggcca gttgtggat
11701 gtggtgctg cgtcgtccg cgaccagcc gttcagccg tcgcgcgat cagcactac
11761 cgcgaaagcc agatcctga cggcgccggg ggcgacgcg tggcggaatt ctgcaacgac
25 11821 gacgtaccg catggtggc cggggcattc cagcccgctg gtgcagcgga caagggcct
11881 gccgaacagc agtggcgga atgggaactg gaactggtc ccacggatgg gaccgccat
11941 accaagctac tggaccggct agccaaccg ctgctgatg ccggtgccg acctgccgc
12001 cacggctcca aactggcgcg ggtgctcgt gcgacctc cgggtgagct gccaacggc
12061 ccgcagccg cggcgatcc agtacaccg gcggtgtcc agcaagtga gcagctgtg
30 12121 ctgtgggatc gggccgtgcg ggccgacgc tatgacccg tgcaccagat gcgagtacg
12181 acccgcaaga tccgagctt gctgacggat tccaggagt cgttggcct gaaggaaat
12241 gcgtgggtca tcgatgaact gcgtgagct gccgatgtc tgggcgtagc ccgggacgc
12301 gaggtactg gtgaccgcta ccagcgcaa ctggacgcg tggcgccga gctggtacg
12361 ggccgggtg gcgagcgct gtagacggg gcgcgcggc gataccagac cgggtgcg
35 12421 cgtactga tcgattcg gtcgacgg tacttccgt tctcgacg tctagacg
12481 cttgtgtcc aacgcgcca tgccactt ggggaggaat cggcaccgt aaccatgat

12541 gcggcctacc ggcgagtccg caaagccgca aaagccgcaa agaccgccgg cgaccaggcg
 12601 ggcgaccacc accgcgacga ggcatggcac ctgatccga agcgcgcgaa gcgattacgc
 12661 tacaccgagg cggtactagg ggccgacaat gtgtcacaag aagccaaggt catccagacg
 5 12721 ttgctaggcg atcatcaaga cagcgtggtc agccgggaac atctgatcca gcaggccata
 12781 gccgcgaaca ccgccggcga ggacaccttc acctacggtc tgctctacca acaggaagcc
 12841 gacttgcccg agcgctgccg ggagcagctt gaagccgcgc tgcgaaaact cgacaaggcg
 12901 gtccgcaaag caggggattg agcccgcag ggccggacga gttggcctgt aagccggatt
 12961 ctgttccgcg ccgccacagc caagctaacg gcggcacggc ggccaccatc catctggaca
 10 13021 caccgttacc gggcgccctg agcgccctac ccgcaggctc ggccgagcaa cctcaagcg
 13081 cctgcgccgc cgcactttcg gtgcggcctt ctggccttg ctccgggtgg ggtttgccta
 13141 gccaccccgg tcaccgggaa tgctggcgcg ctctaccgc accgtttcac ccttgcacc
 13201 acgaggatgg cggctgtttt tctgtggcac ttcccgcga gtcacctcgg attgccgta
 13261 gcaatcacc tgctctgtga agtccggact ttctcgact cgacgtgaa cctcgtgaat
 15 13321 ccacacaagc cctacgcgag ccgcggccgc ccagccaact catccgcgac gaccacgcta
 13381 ccccgctggg cgggtgcgcg gccagtgtga ccgctggacg acacggctag tcggacagcc
 13441 gatccggcgg gcagtcctta tcgtggactg gtgacacggt gggacaaaac cgtcgactcc
 13501 ggcgactggg acgccatcgc tgcgaggctc agcgagtacg gtggcgact gctacctcgg
 13561 ctgatcacc ccggcgaggc cggccggctg cgcaagctgt acggcgacga cggcctgttt
 20 13621 cgctcgacgg tcgatatggc atccaagcgg tacggcgccg ggcagtatcg atatttccat
 13681 gcccctatc ccgagtgatc gagcgtctca agcaggcgct gtatccaaa ctgctgccga
 13741 tagcgcgcaa ctgtggggcc aaactgggccc gggaggcgcc ctggccagac agccttgatg
 13801 actggttggc gagctgtcat gccgccggcc aaaccgac caccgcgtg atgttgaagt
 13861 acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgttccgc
 25 13921 tgcaggtggg gatcaacctg agcgatccgg aaaccgacta caccggcggc gatttctgc
 13981 ttgtcgaaca gcggcctcgc gcccaatccc ggggtaccgc aatgcaact ccgcagggac
 14041 atggttatgt gttcacgacc cgtgatcggc cgggtcggac tagccgtggc tggtcggcat
 14101 ctccagtgcg ccatgggctt tcgactatc gttccggcga acgctatgcc atggggctga
 14161 tctttcacga cgcagcctga ttgcacgcca tctatagata gcctgtctga ttcaccaatc
 30 14221 gcaccgacga tgccccatcg gcgtagaact cggcgatgct cagcgatgcc agatcaagat
 14281 gcaaccgata taggacgccc gacccggcat ccaacgccag ccgcaacaac attttgatcg
 14341 gcgtgacatg tgacaccacc agcaccgtcg cgccttcgta gccaacgatg atccgatcac
 14401 gtccccgccg aaccgcgcgc agcacgtcgt cgaagcttc cccaccggg ggcgtgatcg
 14461 tgggttctct cagccagcga cggtcagct cgggatcgcg ttctgcggcc tccgcgaacg
 35 14521 tcagcccctc ccaggcgccg aagtcggtct cgaccaggtc gtcacgacg accacgtcca
 14581 gggccagggc tctggcggcg gtcaccgcgg tctcgtaacg ccgctgtagc ggccaggaga

14641 ccaccgcagc gatcccgccg cgccgcgcca gatacccgcc cgccgcacca acctggcgcc
14701 accccacctc gttcaacccc gggttgccgc gcccgaata gcggcggtgc tccgacagct
14761 ccgtctgcc gtggcgcaac aaaagtagtc gggtgggtgt accgcggcg ccggtccagc
5 14821 cgggagatgt cgggtactcg gtcgcaacga tttggcagg atccgcatcc gccgcagccg
14881 attgcgggc ggcgtccatc gcgtcattgg ccaaccggtc tgcatacgtg ttccgggcac
14941 gcggaaccca ctgtagttg atcctgcgaa actgggacgc caacgcctga gcctggacat
15001 agagcttcag cagatccggg tgcttgacct tccaccgcc ggacatctgc tccaccacca
15061 gcttgagtc catcagcacc gcggcctcgg tggcacctag ttacggcg tctccaaac
10 15121 cggctatcag gccgcggtat tggcgacgt tgttcgtgc ccggccgac gcctgcttg
15181 actcgccag cacggtggag tgatcgccgg tccacaccac cgcgccgtat ccggccggtc
15241 cgggattgcc ccgcatccg ccgtcggtt cgtgacaac ttactcct caaatcctc
15301 gagccgcaac aagatcgtc cgcatccgg gcagcgacc acttcacct cggcgccgc
15361 cgagatctgg gccagctgc cgcggccgat ctgatccgg caggcaccac atcgatgacc
15421 ttgcaaccgc ccggcccctg gccgcctcc ggcccgtgt cttcgtaga gcccgcgaag
15 15481 ctgggatca agtgcgccc tcagcatgc gcgttcgat gaatgtggt gccgggctg
15541 gtcgatttcg gcaagtgcct cgtccaaagc ctgctggcg gcggccaggt cggcccga
15601 cgcttgagc gccgcgact cggcggtctg ttgagcctgc agtcctcgc ggcgtccag
15661 cacctccagc agggcatctt ccaactggc ttgacggcgt tgcaagctgt cgagctcgtg
20 15721 ctgcagatca gccaattgt tggcgtcgt tgcaccgaa gtgagcaac accggtccc
15781 gtcgccagc ttacgcacc catcgtctc cgactcaaa cgcgacacct ggccgtcaa
15841 gtctccgcc gcgattcga gggccgcat cctgtcgtt gcggcggtt gctcgccgtg
15901 cacctgctgg taagccgcc gctcggcag atgggtagcc cgatgcgca tccgggtcag
15961 ctacgatcc agcttcgcca attcagtag cgaccgttc tgtgccactc cggcttcat
25 16021 gcctgatctc tccagttc gtatcgagg ttccacgggt cggtcagat ggtgcacaca
16081 cgcaccggca gcgacgccc gaaatgagac cgcaacact cggcggcctg gccgcaccac
16141 ggggaattcg ttgccaatg cgcgacgtc atcaggcca cttgcaagc tcggcaatgc
16201 tcgtcggtg gatgatgct cagatcgcc gtaacgtac ctgacgctc cgcggcgcc
16261 acggtggcaa gcaacgagtc ccggcgccg ccgcagacc gcaccgcga caccagcagg
30 16321 tcgggatccc cggcggcgcg cacaccggtc gcagtcggcg gcaacgcggc ctccagacgg
16381 gcaacaaagg tgcgcagcg ttgggtttt ggcagtctc caatccggc taaccgctg
16441 ccgaccggcg gtggtaccg cggaagatg tcgaatccg gtcctcgtg aggtgctgcg
16501 gcgcgcatc ccgcaacac ctggcgcg gctcgtcgg gtgcgacgc ctgaccgg
16561 tcctggcca ccgttcgac ggtaccgac ctgcctatg cggcgacgc ccgtcgtg
35 16621 gccaggaact gcccggtacc cgcgacctc cagctcagt gcgagtagc gccgatatg
16681 ccggcaccg cctcaaagac cgtgcccgc accgcctct agttctcgc cggcacatag

16741 atgaccact tgcgagatc ggccgctccg ggcaccgggt cgagaacggc gtcgacggtc
16801 agaccaacag cgtgtgccag cgcgtcggac acaccggcg acgccgagtc ggcgttggtg
16861 tgcgcggtaa acaacgagcg accggtccgg atcaggcgggt gcaccagcac accctttggc
5 16921 gtgttgccg cgaccgtatc gacccacgc agtaacaac ggtggtgcac caatagcagt
16981 ccggcctggg gaacctggtc caccaccgcc ggcgtcgcgt ccaccgcaac ggtcaccgaa
17041 tccaccacgt cgtcggggtc gccgcacacc agaccaccg aatccacga ctgggcaagc
17101 cgcggcgggt aggcctggtc cagcacgtcg atgacatcgg ccagccgcac actcatcggc
17161 gtcctccacg ctttcccac tcggcgatcg ccgccaccag cacgggccac tccgggcgca
10 17221 ccgccgcccg caggtaccgc gctccaggc cgacgaagg gtcaccgcgg cgcaccgcaa
17281 ttctttgct ctgcaatag ttctgaatc cgtcagcatc ggcgatgtg aacagtacga
17341 aaggggccgc accatcgacc acctcggcac ccaccgatct cagtccggcc accatctccg
17401 cgcgcagcgc cgtcaaccgc accgcacatc cgtcggcagc ggcgaccgcc cggggggcgc
17461 agcaagcagc gatggcgtc agttgcaatg ttccaacgg ccagtgcgt cgtgcacgg
15 17521 tcaaccgagc cagcacgtct ggcgagccga gcgcgtagcc caccgcaat ccggccagcg
17581 accacgttt cgtcaagcta cggagcacca gcacatcggg cagcgagtca tcggccaacg
17641 attcgggtc gccgggaacc caatcagcga acgctcgtc gaccaccagg atgcgtccc
17701 gccggcgtaa ctcgagcagc tgctcggga ggtgcagcac cgagggtggg ttggtcggat
17761 taccacgac gacaaggctg gcgtcgtcag gcacgtcgc ggtgtccagc acgaacggcg
20 17821 gcttaggac aacatggtgc gccgtgattc cggcagcgt caaggctatg gccggctcgg
17881 tgaacgcggg cagcagcatt gctcccgcga ccggacttag gttgtcagc aatgcgaatc
17941 cctccgccgc cccgacgagc gggagcactt cgtcacgggt tctccatga cgttcagcga
18001 ccgcgtcttg ccccgggtgc acatcgtcgg tgctcggata gcgggccagc tccggcagca
18061 gcgcggcgag ctgccggacc aaccattccg ggggccggtc atggcggacg ttgacggcga
25 18121 agtcacgac gccgggcgcg acatcctgat caccgtggta gcgcgccgcg gcaagcgggc
18181 tagtgtctag actcgccaca gcgtcaaca gtagtgggccc ggtgtgcggg ccaagaatcc
18241 agagcaccgc cgacgcgtt tctacgcggc gacaaccgcg acatcacagg cagctaacag
18301 ggcgtcggcg gtgatgatc tcaggccaag cagctgtgcc tgggcgatga gcacacggtc
18361 gaatggatgt cgatggtgat ccggaagctc tcgggtgcgc agtgtgtgcg tggtaactg
30 18421 acagcggcga cgtcccgcag cggcgcatc gatcgggcac gtaagaagcc gatggctcgg
18481 gcggcgggag cttgccgagg cggtagtga tcgcgatc ccaggcactg gcggccgaca
18541 agagaatgct gttgcggacg tcctgaacaa tcgccgtgt ttcgtgacg gcacccgag
18601 ccaaactggg gtgtcgatga ggtagcgtt caccgtgaa agcgttcgag cacgtcgtt
18661 gacaacggag cgtccaaatc gtcgggcacg cggtagcgc catggtcaat gcctaaccg
35 18721 cgagtctcat gaggatcag cggcacaagc ttgctaccg gtcgccgcg gcgggcaatc
18781 tcaacctctg cccgccgtag acgagccgca gcagctcggc caggcgtgtc ttcgcctcgt

18841 gaacgccgac ccgttcgca ggcccccaga ctttcgcgtc gaccacctgc tcaccaaact
18901 tcgcgatcat cgctgatac cacagcgcca acgggtagcg gttgtccaa ccgttcgtc
18961 aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaattgccc gcctcctcca
5 19021 cgcgccgccc caggcgcgcg atcgtcgccc ggtgaatcgc cgcagctggt gatcttcgat
19081 ctggacggca cgtgaccga ctcgcgcgcg ggaatcgtat ccagcttccg acacgcgctc
19141 aaccacatcg gtgccccagt acccgaaggc gacctggcca ctacatcgt cggccccccc
19201 atgcatgaga cgctcgcgcg catggggctc gggaatccg ccgaggaggc gatcgtagcc
19261 taccggggccg actacagcgc ccgcggttgg gcgatgaaca gctgttoga cgggatcggg
10 19321 ccgtctgtgg ccgacctgcg caccgccgtt gtccggctgg ccgtgccac ctccaaggca
19381 gagccgaccg caggcggaat cctgcgccac ttccgaattg agcagcactt cgaggctatc
19441 gcgggcgcca gcaccgatgg ctgcgaggc agcaaggcgc acgtgctggc ccaacgcgtc
19501 gcgcagctgc ggccgtacc cgagcgggtg gtgatggtcg gcgaccgcag ccacgacgtc
19561 gacggggcgcg ccgcgcacgg catcgacacg gtggtggtcg gctggggcta cggcgcgccc
15 19621 gactttatcg acaagacctc caccaccgtc gtgacgcgtc ccgccacgat tgacgagctg
19681 agggaggcgc tagtgtctg atccgtgca cgtcacattc gttgtacgg gcaacatctg
19741 ccggtcgcca atggccgaga agatgttcgc ccaacagctt cggcaccgtg gcctgggtga
19801 cgcggtgcca gtgaccagt cgggcaccgg gaactggcat gtaggcagt ggcgccgaca
19861 gcgggcgccc ggggtgttc gagccacgg ctaccctacc gaccaccggg ccgcacaagt
20 19921 cggcaccgaa cactggcgg cagacctgtt ggtggccttg gaccgcaacc acgtcggct
19981 gttcgggcag ctggcgctc aagccgccc ggtacggatg ctgcggtcat tcgaccacg
20041 ctcggaacc catgcgtcg atgtcgagga tccctactat ggcatcact ccgacttga
20101 ggaggtcttc gccgtcatc aatccgccct gcccgccctg cagactggg tcgacgaacg
20161 tctcgcgcg aacggaccga gttgatgcc cgctagcgt tctgtctcg gcccggtgg
25 20221 ctggcgttgg cctgtgctg gtcgcgttc acctacctgt gcttacggt gtcgcgccg
20281 tggcagctgg gcaagaatgc caaacgtca cgagagaacc agcagatcag gtattccctc
20341 gacacccgc cgttccgct gaaaacctt ctaccacagc aggatcgtc ggcgccggac
20401 gcgcagtggc gccgggtgac ggcaaccgga cagtacctc cggacgtgca ggtgctggc
20461 cgactgcgcg tggaggagg ggaccaggcg ttgaggtgt tggccccatt cgtggtcgc
30 20521 ggcggacca cctgctggt cgacctgga tacgtcggc ccaggtggg ctgcacgta
20581 ccaccgatcc ccgcctgcc ggtgcagacg gtgacatca ccgcgcggct gcgtgactc
20641 gaaccgagcg tggcgggcaa agaccattc gtcagagacg gcttccagca ggtgtattc
20701 atcaataccg gacaggtcgc cgcgtgacc ggagtccagc tggctgggtc ctatctgcag
20761 ttgatcgaag accaaccgg cgggctcggc gtgctcggcg ttccgatct agatccggg
35 20821 ccgttctgt cctatggcat ccaatggatc tcttcggca ttctggcacc gatcggtt
20881 ggctatttcg cctacgccga gatccggcg cgccggggg aaaaagcggg gtcgccacca

20941 ccggacaagc caatgacggt cgagcagaaa ctgctgacc gctacggccg ccggcggttaa
 21001 accaacaatca cggccaatac cgcagcccc gcctggacca cccgcgacag caccacggcg
 21061 cggcgcagat cggccacctt gggcgaccgg ccgtcgccca aggtgggccc gatctgcaac
 5 21121 tcatggtggt accgggtggg cccaccagc cgcacgtcaa ggcggccagc aaacgcggcc
 21181 tcgacgacac cggcgttggg gctgggatgg cggcgggcgt cgcgccgcca ggcccgctacc
 21241 gcaccgcggg gcgaccacc gaccaccggc gcgcagatca ccaccagcac cgcgctgcgc
 21301 cgtgcgcaa catagtggc ccagtcatcc aatcgtgctg cagcccaacc gaatcggaga
 21361 taacgcggcg agcggtagcc gatcatcgag tccagggtgt tgatggcacg atatccagc
 10 21421 accgcaggca cgcgctcga agccggccac agcagcggca ccacctgggc gtcggcggtg
 21481 tttcggcca ccgactccag cgcggcacgc gtcaggcccg ggccggccag ctgggcccgg
 21541 tcacggccgc acagcgacgg cagcagccgt cgcggccct cgacatcgtc gcgtccaac
 21601 aggtccgata tctggcgccc ggtgcgcgc agcgaagttc cggcagcgc tgcgcagggtg
 21661 gccgtcgcgg tggccggccac gggccaggac ctgccgggta gccgtgcag tgcgcggcg
 15 21721 agcaagccca ccgcggccac cagcaggccc acgtgtaccg caccggcgac ccggccgtca
 21781 cggtaggtga tctgctccag ctggcgccc gcccgaccga acaggggccac cggatgacct
 21841 cgtttggggt cggcgaacac gacgtcgagc aggcagccga tcagcacgcc gacggccctg
 21901 gtctgccagg tcgatgcaaa cactccggca gcgtcgaca cgtggtctac gctcagtat
 21961 ttatgacctc atacggcagc tatccacgat gaagcggcca gctaccggg ttgccgacct
 20 22021 gttgaacccg gcggcaatgt tgtgcccgc agcgaatgtc atcatgcagc tggcagtgcc
 22081 ggggtgctggg tatggcgtgc tggaaagccc ggtggacagc ggcaactct acaagcatcc
 22141 gttcaagcgg gcccgacca ccggcaccta cctggcggtg gcgaccatcg ggacggaate
 22201 cgaccgagcg ctgatccggg gtgcccgtga cgtcgccac cggcaggttc ggtcgacggc
 22261 ctgagccca gtgtctata acgccttca cccgaagtgc cagctgtggg tggcggcgtg
 25 22321 tctgtaccgc tacttctggt accagcacga gtttctgtac ggccactcg aagatgccac
 22381 cgccgacgcc gtctaccaag acgccaacg gttagggacc acgtgcagg tgcggaggg
 22441 gatgtggccc ccggaccggg tcgcttca cgagtactgg aagcgctcgc ttgatgggt
 22501 gcagatcgac gcgcccgtgc gcgagcatct tcggggggtg gcctcgtag cgtttctccc
 22561 gtggccgttg cgcgcggttg ccggccgtt caacctgtt gcgacgacgg gattcttggc
 30 22621 accggagttc cgcgcgatga tgcagctgga gtggtcacag gccagcagc gtcgcttga
 22681 gtggttactt tccgtgctac ggtagccga ccggctgatt ccgcatcggg cctggatctt
 22741 cgtttaccag ctttactgt gggacatgcg gtttcgcgc cgacacggcc gccgaatcgt
 22801 ctgatagagc ccggccgagt gtgagcctga cagccgaca ccggcgcggt gtgtcgcgtc
 22861 gccaggttca cgtcggcga tctagaccg ccgaaaacct acttctgggt tgcctccga
 35 22921 atcaacgtgc tgatctgctc gagcagctca cgcatacgg cgcgcatcgc atccaccg
 22981 gcatacaggt cggccttggg ccggcgacg tggccgacg tcattggccg caccggcggt

23041 gctgtctgtc ggcgcgcgt gtcgcttga aaccaggctc gtcacccac gaccacgaca
 23101 ctgccatata cggcgccccg ccgacaacga agcacagcta gccggtgggc gcggacggga
 23161 tcgaaccgcc gaccgtggt gtgtaaaacc agagctctac cgctgagta cgcgccccatg
 5 23221 accgccgcag gctacacgcc ttgcggccaa gcacccaaaa ccttaggccg taagcgccgc
 23281 cagagcgtcg gtccacagcc gctgatcgcg aactcaccc ggctgcttca tctcggcgaa
 23341 ccgaatgac cctgaccgat cgaccacaaa ggtgccccgg ttagcgatgc cggcctgtc
 23401 gttgaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg
 23461 aaacgtgaat ccgctctcg tcgccagat ctgtgagtg ggtggcgggc ccaccgaaat
 10 23521 cgctagcgcg gcgtgtcgt cgttctcaa ctcgggcagg tgatcacga actggtccag
 23581 ctgcacctgg cagatgcccg tgaacgcaa cggaaagaac accaacagca cgttcttgc
 23641 acccggtag ccgcgcaggg tgacaagctg ctgattctgg tcgcgaacg tgaagtcagg
 23701 ggcggtggct ccgacgttca gcatcagcg tcgccagccc gcgatttcgg ctgtaccaat
 23761 ctgctggcgc tccagtgcc cagattgacc gacgaggtcg gcatcagccc agctgtgggc
 15 23821 gccgcctcgg caatctcggc gggcaataca tggccgggct ggccggtctt ggcgctcacc
 23881 acccaaatca caccgtcctc ggcgagcggg ccgatcgcat ccatcagggt gtccacaaa
 23941 tcgccgtcgc catcacgcca ccacaacagg acgacatcga tgacctcgc ggtgtcttca
 24001 tcgagcaact ctccccgca cgcttctcg atggccgcgc ggatgtcgc gtcggtgtct
 24061 tcgtcccagc cccattcctg gataagttgg tctcgttga tgcccaattt gcgggcgtag
 20 24121 ttcgaggcgt gatccggcg gaccacgtg gaacctcctt cagtctcgc gggccatgtg
 24181 cacaccgtcg cgatggcat tatcgtcga cagccagaac cggccaccc gccgcctca
 24241 gaaggcggcc acgcacattg tcaatgcctt tgtcttggtg tcgttgagcc gatcaaccg
 24301 ccggttgaat tccgtgtcg acgctgcgc accgatggca ttgccaccg cgcgggccc
 24361 gtcgacatat gcgttgagcg catccccag ttgcgcggac agcgcggcgc tcagactgcc
 24421 tgagaccgtc gaggcactgt tgttgagcg gtcgatggcc ggaccttcgg tcggcccgt
 25 24481 gttgcggccc tgattgaac cgccacgta ggcgttcacc ttgtgatgg ctccttgc
 24541 ggtggccgcc agcgcgtcac acgaggtcg aatgccttg gtcgtcagcg attgttggc
 24601 ctgcgactcc cgatgtctc acgtcgccgc cgaagccgac accgacgcgg acaccgacga
 24661 gcggtaggcc ggtgcgacgt tgggtcggg catggccgta ccgtcgtga cagtgttaca
 30 24721 tccgacgatc cccatcagca gcagcgcgat gcagccgagc gccagggcgc ctcgctggg
 24781 gagtcccc ccgtgcctgc gaggcacgc gcgccatccg atgagcacgg catgtgaggt
 24841 tacctggtcg cagcgcgacc gcgtggccg tgggtgtcg cgcacccga gaaccgagc
 24901 gagtgcggct atccgccgc gacccgggtg cggcacgata gggggacgac catctaaaca
 24961 gcacgcaagc ggaagccgc cacctacagg agtagtgcgt tgaccaccga ttccgccgc
 35 25021 cacgatctgg cccaaaact aaacagcgca agcgaacccg accgagttcg ggtgatccg
 25081 gaggggtgtg cgtcgtattt gcccgacatt gatcccgagg agacctcgga gtggctggag

25141 tcctttgaca cgtgctgca acgtgcggc cgtcgcggg cccgctacct gatgttgcg
 25201 ctgctagagc gggccggcga gcagcgggtg gccatcccg cattgacgtc taccgactat
 25261 gtcaacacca tcccgaccga gctggagccg tggttccccg gcgacgaaga cgtcgaacgt
 5 25321 cgttatcgag cgtggatcag atggaatcg gccatcatgg tgcaccgtgc gcaacgaccg
 25381 ggtgtgggcg tgggtggcca tatctgacc tacgcgtcgt ccgcggcgt ctatgaggtc
 25441 ggtttcaacc acttcttcg cggcaagtcg caccggggcg gcggcgatca ggtgttcac
 25501 cagggccacg ctccccggg aatctacgcg cgcgccttc tcgaaggcg gttgaccgc
 25561 gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcggtt gccgtcctat
 10 25621 ccgcaccgcg ggctcatgcc cgacttctgg gaattccca ccgtgtcgat gggtttggc
 25681 ccgctcaacg ccatctacca ggcacggtc aaccactatc tgcattgacc cggtatcaaa
 25741 gacacctcg atcaaacgt gtggtgttt ttggcgacg gcgagatgga cgaaccgag
 25801 agccgtggcg tggccacgt cggcgcgctg gaaggcttg acaactgac ctctgtgac
 25861 aactgcaatc tgcagcgact cgacggccc gtgcgggca acggcaagat catccaggag
 15 25921 ctggagtcgt tcttcgcg tgcggctgg aacgtcatca aggtggtgtg gggccgcgaa
 25981 tgggatgccc tgctgcacgc cgaccgcgac ggtgcgtgg tgaattaat gaatacaaa
 26041 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccactc
 26101 ttggccgcg acccagcac caaggcgctg gtggagaaca tgagcgacca ggatatctg
 26161 aacctcaaac gggcgggcca cgattaccg aaggtttacg ccgctaccg cgcgcgctc
 20 26221 gaccacaagg gacagccgac ggtgatcctg gccaagacca tcaaaggcta cgcgtggcg
 26281 aagcatttcg aaggacgcaa tgccaccac cagatgaaa aactgaccct ggaagacct
 26341 aaggagttc gtagacgca gcgattccg gtcagcgacg ccagcttga agagaatccg
 26401 tacctgccg cctactacca cccggcctc aacgcccgg agattcgta catgctgac
 26461 cggcgccggg ccctcggggg cttgttccc gagcgagga ccaagtcaa agcgtgacc
 25 26521 ctgccgggtc gcgacatcta cgcgcgctg aaaaagggt ctgggcacca ggaggtggc
 26581 accaccatgg cgacggtgcg caggttcaa gaagtgtgc gcgacaagca gatcggggc
 26641 cggatagtcc cgatcattcc cgacgaggcc cgcaccttc ggatggactc ctggttccc
 26701 tcgctaaaga tctataaccg caatggccag ctgtataccg cggttgacgc cgacctgat
 26761 ctggcctaca aggagagcga agtcgggcag atcctgcacg agggcatcaa cgaagccggg
 30 26821 tcggtgggct cgttcacgc gcccggcacc tcgtatgca gcacaacga accgatgatc
 26881 cccatttaca tctctactc gatgttcggc ttccagcga ccggcgatag cttctgggccc
 26941 gcggccgacc agatggctcg aggttcgtg ctcggggcca ccggggcg caccacctg
 27001 accggtgagg gctgcaaca ccggacggt cactcgttc tgctggcg caccaaccg
 27061 gcggtggtg cctacgacc gcccttcgcc tacgaaatc cctacatcgt ggaaagcgga
 35 27121 ctggccagga tgcggggga gaaccggag aacatctct tctacatcac cgtctacaac
 27181 gagccgtacg tgcagccgcc ggagccggag aacttcgatc ccgaggcg gctcggggg

27241 atctaccgct atcacgcggc caccgagcaa cgcaccaaca aggcgcagat cctggcctcc
 27301 ggggtagcga tgcccgggc gctgcgggca gcacagatgc tggccgccga gtgggatgtc
 27361 gccgccgacg tgtggtcggg gaccagtggg ggcgagctaa accgcgacgg ggtggccatc
 5 27421 gagaccgaga agctccgcca ccccgatcgg ccggcgggcg tgccctacgt gacgagagcg
 27481 ctggagaatg ctcggggccc ggtgatcgcg gtgtcggact ggatgcgcgc ggtccccgag
 27541 cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcaccgacgg gttcggcttt
 27601 tccgacactc ggcccggcg tcgccgtac ttcaacaccg acgccgaatc ccaggtggtc
 27661 gcggttttgg aggcgttggc gggcgacggc gagatcgacc catcgggtcc ggtcgcggcc
 10 27721 gcccgcaggt accggatcga cgacgtggcg gctgcggccg agcagaccac ggatcccggt
 27781 cccggggcct aacgccggcg agccgaccgc ctttggccga atctccaga aatctggcgt
 27841 agcttttagg agtgaacgac aatcagttgg ctccagttgc ccgccgagg tcgccgctcg
 27901 aactgctgga cactgtgccc gattcgtgc tcggcggtt gaagcagtac tcgggcccgc
 27961 tggccaccga ggcagtttcg gccatgcaag aacggttgc gttcttcgc gacctagaag
 15 28021 cgtcccagcg cgccagcgtg gcgctggtgg tgcagacggc cgtggtcaac ttctgcaat
 28081 ggatgcacga ccgcacagt gacgtcggct ataccgcga ggcattcgag ctggtgcccc
 28141 aggatctgac gcgacggatc gcgctgcgcc agaccgtgga catggtgcgg gtcaccatgg
 28201 agttcttcga agaagtcgtg cccctgctcg cccgttcga agagcagttg accgccctca
 28261 cgggtggcat ttgaaatac agccgcgacc tggcattcac cgccgccacg gcctacgccg
 20 28321 atcgggccga ggcacgagcg acctgggaca gccggatgga ggcagcgtg gtggacgagg
 28381 tggtagcgcg cgacaccggg cccgagctgc tgtccgggc ggccgcgctg aattgggaca
 28441 ccaccgcgcc ggcgaccgta ctggtgggaa ctccggcgcc cgggtccaat ggctcaaca
 28501 gcgacggcga cagcgagcgg gccagccagg atgtccgca caccgaggct cgccacggcg
 28561 gcgctgcgt gaccgacgtg cacggcacct ggctggtggc gatcgtctcc ggccagctgt
 25 28621 cgccaaccga gaagttctc aaagacctgc tggcagcatt cgccgacgcc ccggtggtca
 28681 tcggcccccac ggcgcccatg ctgaccggcg cgaccgcag cgctagcgag gcgatctccg
 28741 ggatgaacgc cgtcgccggc tggcgcgag cgcccgggcc cgtgctggct agggaaactt
 28801 tgcccgaacg gcacctgatg ggcgacgcct cggcgatcgt ggccctgcat accgacgtga
 28861 tgcggccccct agccgatgcc ggaccgacgc tcacgagac gctagacgca tatctggatt
 30 28921 gtggcgggcg gattgaagct tgtccagaa agttgttcgt tcaccaaac acagtgcggt
 28981 accggctcaa gcggatcacc gacttcaccg ggcgcgatcc caccagcca cgcgatgcct
 29041 atgtccttcg ggtggcgggc accgtgggtc aactcaacta tccgacgcc cactgaagca
 29101 tcgacagcaa tgccgtgta tagattccct cgccggtcag aggggggtcca gcagggggcc
 29161 cggaaagata ccagggggcg cgtcggacgg aaagtatcc agacaacagg tcgcggggacg
 35 29221 atctcaaaaa catagcttac agggccgttt tgttggttat atacaaaaac ctaagacgag
 29281 gttcataatc tgttacaccg cgaaaaccg tcttcacagt gttctcttag acacgtgatt

29341 gcgttgctcg caccgggaca gggttcgcaa accgagggaa tgtgtcgcc gtggcttcag
 29401 ctgcccggcg cagcggacca gatcgggcg tggtcgaaag ccgctgatct agatcttgcc
 29461 cggctgggca ccaccgctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg
 5 29521 atcgtcgccg cgactctgct ggcccaccag gaactggcgc gccgatgcgt gctcgccggc
 29581 aaggacgtca tcgtggccgg cactccgctc ggcgaaatcg cggcctacgc aatcgccggg
 29641 gtgatagccg ccgacgacgc cgtcgcgctg gccgccaccc gggcgccga gatggccaag
 29701 gcctgcgcca ccgagccgac cggcatgtct gcggtgctcg gcggcgacga gaccgaggtg
 29761 ctgagtcgcc tcgagcagct cgacttggtc cgggcaaacc gcaacgccgc cggccagatc
 10 29821 gtcgtgccc gccggctgac cgcgttgag aagctcgccg aagaccgcc ggccaaggcg
 29881 cgggtgctg cactgggtgt cggcgagcg ttccacacc agttcatggc gcccgactt
 29941 gacggcttg cggcgccgc ggccaacatc gcaaccgccg accccaccgc cacgtgctg
 30001 tccaaccgcg acgggaagcc ggtgacatcc gggccgcgg cgatggacac cctggtctcc
 30061 cagtcaccc aaccggtgc atgggacctg tgcaccgca cgtgcgca acacacagt
 15 30121 acggcgatcg tggagttccc ccccgggc acgcttagcg gtatcgcaa acggaactt
 30181 cggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta
 30241 taaccgcgga ctggccaga acaaccacat acccgctcgt tcgatttga cacaacatat
 30301 tacgaaggga agcatgctgt gcctgtact caggaagaaa tcattgccg tatcgccgag
 30361 atcatgaag agtaaccgg tatcgagccg tccgagatca cccggagaa gtcgttcgtc
 20 30421 gacgacctg acatcgactc gctgtgatg gtcgagatc ccgtgcagac cgaggacaag
 30481 tacggcgta agatccccga cgaggacctc gccggtctgc gtaccgtcg tgacgtgtc
 30541 gcctacatcc agaagctcga ggaagaaaac ccggaggcgg ctcaggcggt gcgcgcgaag
 30601 attgagtcgg agaaccgga tgcggttgc aacgttcagg cgaggctga ggccgagtc
 30661 aagttagtca gccttcacc gctaattggc gttccccag cgttgtgtg accgccgtca
 25 30721 cagcgacgac gtcgatctc cggacatcg agagcacgtg gaagggtctg ttggccggcg
 30781 agagcggcat ccacgcactc gaagacgagt tcgtaccaa gtgggatcta gcggtcaaga
 30841 tcggcggtca cctcaaggat ccggtcgaca gccacatggg ccgactcgac atgcgacgca
 30901 tgcgtacgt ccagcggatg ggcaagttgc tggcgggaca gctatgggag tccgccggca
 30961 gcccgaggt cgatccagac cggttcgcc ttgtgtcgg caccggtcta ggtggagccg
 30 31021 agaggattgt cgagagctac gacctgatga atcgggcgg ccccggaag gtgtccccgc
 31081 tggccgttca gatgatcatg cccaacggtg ccgcgcggt gatcggtctg cagcttggg
 31141 cccgcgccgg ggtgatgacc ccggtgtcgg cctgtctgc gggctcgga gcgatcgccc
 31201 acgcgtggcg tcagatcgt atgggcgacg ccgacgtcgc cgtctcggc ggtgtcgaag
 31261 gaccatcga ggcgtgccc atcgcgcggt tctcatgat gcgggcatg tcgaccgca
 35 31321 acgacgagcc tgagcgggcc tccggccgt tcgacaagga ccgcgacggc ttgtgttcg
 31381 gcgaggccgg tgcgtgatg ctcacgaga cggaggagca cgccaaagcc cgtggcgcca

31441 agccgttggc ccgattgctg ggtgccggtg tcacctcgga cgcctttcat atggtggcgc
 31501 ccgcggccga tgggttctgt gccggtaggc cgtgactcg ctgctggag ctggccgggt
 31561 tgtgccggc ggacatcgac cagtcacg cgcacggcac ggcgacgct atcggcgacg
 5 31621 ccgcggaggc caacgccatc cgcgtcgccg gttgtgatca ggccgcgggtg tacgcgcga
 31681 agtctgcgt gggccactcg atcggcgcgg tcggtgcgt cgagtcgggtg ctacgggtgc
 31741 tgacgtcg cgcggcgct atcccgcga ccctgaacta cgagacaccc gatcccgaga
 31801 tcgacctga cgtcgtgcc ggcgaaccgc gctatggcga ttaccgtac gcagtcaaca
 31861 actcgttcgg gttcgccgc cacaatgtg cgcttcctt cgggcgttac tgaagcacga
 10 31921 catcgccggg cgcgaggccc gaggtggggg tcccccgct tgcgggggcg agtcggaccg
 31981 atatggaagg aacgttcga agaccaatga cggagctggt taccgggaaa gcctttcct
 32041 acgtagtcgt caccggcatc gccatgacga ccgcgtcgc gaccgacgc gagactacgt
 32101 ggaagtgtt gctggaccgc caaagcggga tcgtacgt cgatgacca ttcgtcagg
 32161 agttcgacct gccagttgc atcggcggac atctgctga ggaattcgac caccagctga
 15 32221 cgcggatcga actgcgccg atgggatacc tgcagcggat gtccaccgtg ctgagccggc
 32281 gcctgtggga aaatgccgc tcaccgagg tggacacaa tcgattgatg gtgtccatc
 32341 gcaccggcct gggttcgcc gaggaactgg tcttcagta cgacgatatg cgcgtcgcg
 32401 gaatgaaggc ggtctgccg ctgaccgtgc agaagtacat gccaacggg gccgccgcg
 32461 cggtcgggtt ggaacggcac gccaaaggcg gggatgatg gccggtatc gcgtgcgat
 20 32521 ccggcgccga ggccatgcc cgtgcgtggc agcagattgt gctgggagag gccgatgcc
 32581 ccatctcgg cgcggtggag accaggatc aagcgggtcc catcgccggg ttcgtcaga
 32641 tgcgcatcgt gatgtccacc aacaacgacg accccgccg tgcagccgc ccattcgaca
 32701 gggaccgcga cggctttgt ttcggcgagg gcggcgccct tctgtgatc gagaccgagg
 32761 agcacgcaa ggcacgtggc gccaatcc tggcccgat catggcgcc agcatcacct
 25 32821 ccgatggct ccacatggtg gcccggacc ccaacgggga acgcgccggg catgcgatta
 32881 cgcggcgat tcagctggcg ggcctcgcc ccggcgacat cgaccacgtc aatgcgcac
 32941 ccaccggcac ccaggtcggc gacctggccg aaggcagggc catcaacaac gccttggcg
 33001 gcaaccgacc ggcggtgtac gccccaaat ctgccctcg ccactcgggt ggcgcggtc
 33061 gcgcggtcga atgatcttg acggtgctc cgttcgcga tcaggtgatc ccgccgac
 30 33121 tgaatctgt aaacctgat ccgagatc atttgacgt ggtggcgggt gaaccgcac
 33181 cgggcaatta ccggtatcg atcaataact cgttcggtt cggcgccac aacgtggca
 33241 tcgcttcgg acggtactaa acccagcgt tacgcgacag gagacctcg atgacaatca
 33301 tggccccga ggcggtggc gattcgtcg accccgcga tccgtgttg cggtgagca
 33361 acttctcga cgacggcagc gtggaattgc tgcagagcg tgaccgtcc ggagtgtg
 35 33421 ccgcggcggg caccgtcaac ggtgtgcga ccatcggtt ctgcaccgac ggcaccgtga
 33481 tggcgggcgc catggcgct gaggggtga cgcacatcgt caacgcctac gacactgca

33541 tcgaagacca gagtcccatc gtgggcatct ggcattcggg tgggtcccgg ctggctgaag
 33601 gtgtgcgggc gctgcacgcg gtaggccagg tgttcgaagc catgatccgc gcgtccggct
 33661 acatcccgcga gatctcgggtg gtcgtcgggt tcgccgccgg cggcgccgcc tacggaccgg
 5 33721 cgttgaccga cgtcgtcgtc atggcgccgg aaagccgggt gttcgtcacc gggcccgcac
 33781 tgggtgcgag cgtcaccggc gaggacgtcg acatggcctc gtcgggtggg ccggagaccg
 33841 accacaagaa gtccgggggtg tgccacatcg tcgccgacga cgaactcgat gcctacgacc
 33901 gtgggcgccg gttggtcgga ttgtctgcc agcaggggca ttctgatcgc agcaaggccg
 33961 aggccgggtga caccgacatc caccgcgtgc tgcgggaatc ctgcgcagct gcctacgacg
 10 34021 tgcgtccgat cgtgacggcg atcctcgatg cggacacacc gttcgacgag ttccaggcca
 34081 attgggccc gtcgatgggt gtcgggctgg gtcggctgtc gggtcgcacg gtgggtgtac
 34141 tggccaacaa cccgctacgc ctgggcgggt gcctgaactc cgaagcgca gagaaggcag
 34201 cgcgtttcgt gcggtgtgc gacgcgttc ggattccgct ggtggtggtg gtcgatgtgc
 34261 cgggctatct gcccgtgtc gaccaggagt ggggtggcgt ggtgcgccgt ggcgccaaat
 15 34321 tgctgcacgc gttcggcgag tgcaccgttc cgcgggtcac gctggtcacc cgaagacct
 34381 acggcggggc atacattgcg atgaactccc ggtcgttgaa cgcgaccaag gtgttcgcct
 34441 ggccggacgc cgaggtcgcg gtgatggcg ctaaggcggc gtcggcatc ctgcacaaga
 34501 agaagtggc cgcgctccg gacacgaac gcgaagcgt gcacgaccag ttggccgccg
 34561 agcatgagcg catcgccggc ggggtcgaca gtgcgtgga catcggtgtg gtcgacgaga
 20 34621 agatgaccc ggcgatact cgcagcaagc tcaccgaggc gctggcgag gctccggcac
 34681 ggcgcgcccg ccacaagaac atcccgtgt agttctgacc gcgagcagac gcagaatcgc
 34741 acgcgcgagg tccgcgccgt gcgattctgc gctgctgc cagttatccc cagcgggtgc
 34801 tggtaacgc gaggcgtcc tcgcatgctc ggacggtgcc taccgacgcg ctaacaattc
 34861 tcgagaaggc cggcgggttc gccaccaccg cgcaattgct cacggtcatg acccgccaac
 25 34921 agctcgacgt ccaagtga aaacggcgcc tcgttcgctg ttgtacggg gtctacgcgg
 34981 cacaagacc ggacctgtt ggccgcttg cggctctcga tgtgtcatg ggggggcacg
 35041 ccgtcgctg tctgggcacc gccgccgct gtatggatt cgacacggaa aacaccgtcg
 35101 ctatccatat gtcgatccc ggagtaagga tgcggccac ggtcggctg atgtccacc
 35161 aacgcgtcgg tgcgggctc caacgggtgt caggctgtc cgcgaccgc cccgcatgga
 30 35221 ctgccgtgga ggtcgcacga cagttgcgc gcccgcgggc gctggccacc ctgcacgccg
 35281 cactacggtc aatgcgtgc gtcgcagtg aaattgaaaa cgccgtgct gagcagcgag
 35341 gccgccgagg catcgtcgc gcgcgcgaac tcttaccctt cccgacgga cgcgcggaat
 35401 cggccatgga gagcgaggct cggctcgtca tgatcgacca cgggctgcc ttgccgaac
 35461 ttaataccc gatacacgc caggtggtg aaatgtggc agtcgactc gcctggccc
 35 35521 acatcgctct cgcggccgaa tacgaaagca tcgagtggca cgcgggaccg gcggagatgc
 35581 tgcgcgacaa gacacgctgg gccaaagctc aagagctcg gtggacgatt gtcccattg

35641 tcgtcgacga tgtcagacgc gaacccggcc gcctggcggc ccgcatcgcc cgccacctcg
35701 accgcgcgcg tatggccggc tgaccgtgg tgagcagacg cagagtcgca ctgcggccgg
35761 cgcagtgcga ctctgcgtct gctcgcgtc aacggctgag gaactccta gccacggcga
5 35821 ctacgcgtc gcgatcccg ggaccagac cgatccgggt ccggcggctg aggatatcgt
35881 ccacatccag cgcctcccta tgggtcaccg cgtattcga ctccgcccg gtcacgtcga
35941 tgccgtcggc gaccggctcg gtgggcccgt cacatgtggc ggcggcagcg acgttggccg
36001 cctcgcccc gtaccgcgc accagcgact cgggcaatcc ggcgcccgat ccggggggccg
36061 gccagggtt cgcgggtcg ccgatcagcg gcaggttcg agtgcggcac ttgcggctc
10 36121 gcaggtgtcg cagcgtgatg gcgcgattca gcacatctc tgccatgtag cgtattccg
36181 tcagctgcc gccgaccaca ctgatcagc ccgacggcga ttcaaaaaca gcgtggtcac
36241 gcgaaacgtc ggcggtgcgg ccctggacac cagcaccgcc ggtgtcgatt agcggccgca
36301 atccgcata ggaccgatg acatcctgg tgccgaccgc cgtcccaat gcggtgttca
36361 ccgtatccag caggaacgtg atctctccg aagacggtt ggccacatcg ggaatcgggc
15 36421 cgggtgcgtc ttctcggtc agcccgagat agatccggcc cagctgctcg ggcatggcga
36481 acacgaagcg gttcagtc cccgggatcg gaatggtcag cgcggcagtc ggattggcaa
36541 acgacttcg gtcgaagacc agatgtgtc cgcggctggg gcgtagctc agggacgggt
36601 cgatctacc cgcacacag cccgcccggt tgatgacggc acgcgccgac agcgcaaacg
36661 actgccgggt gcgcggctg gtcaactcca ccgaagtcc ggtgacatt gacgcgcca
20 36721 cgtaagttag gatgcggcg ccgtgtggg ccgcggctcg cgcgacggcc atgaccagcc
36781 gggcgctcgt gatcaattgc ccgtctacg cgagcagacc accgtcgagg ccgtccgcc
36841 gaacggtggg agcaatctc accaccgtg acgcgggat tcggcgcgat cggggcaacg
36901 tcgcccccg cgtaccgct agcaccgca aagcgtcgcc ggccaggaaa ccggcacgca
36961 ccaacgccc cttggttga ccatcgacg gcaacaacgg gaccagtgc ggcatggcat
25 37021 gcacgagatg aggagcgtt cgtgtcatca ggattccg cgtgacggcg ctgcgcccgg
37081 cgatgccac gttgccgtg gccagatagc gcagaccgcc gtgcaccaac ttgagctcc
37141 agcggctggt gccgaacgcc agatcatgct ttccaccaa ggccaccgtc agaccgggg
37201 tggcagcatc taaggcaatg ccaacaccg taatgccgc gcctatcac atgacgtcga
37261 gtgcgccacc gtcggccagt gcggtcaggt cggcggagcg acgcgcccg ttgagtgcag
30 37321 ccgagtggg catcagcaca aatatccgt cagtgcgtg gtaagtccg tggccagcg
37381 ggcggaatc aggatcgaat cgacgatgc cgcggactgg atggtcgact gggcgatcag
37441 caacacatg gtcgccagt gacgagcgt gccggagcg acactgccc accgtgcg
37501 cactgtcagc cggcgggcca accctcgat caggacctgc tggctggtc cgaggcgtc
37561 ggtgatgtac accctggcca gtcggagt catgaccgac atgatcagat cgtaccccg
35 37621 caaccggtc gccaccgca caatctgct taccaacgt tccggctgt cccgctcag
37681 gggcacctcc cgcagcacgt cggcgatatg gctggtcagc atggacgca tgatcgaccg

37741 ggtgtccggc cagcgacggt atacggtcgg gcggctcag cccgcgcgcc gggcgatctc
 37801 ggcaagtgtc acccggcca cgccgtaac gacgacgcag ctgcccgtg cccgcaggat
 37861 acgaccaccg gtatccgcgc ggtcattact cattgacagc atgtgtaata ctgtaacgcg
 5 37921 tgactcaccg cgaggaactc ctccaccga tgaatggga cgcgtgggga gatccccccg
 37981 cggccaagcc actttctgat ggcgtccggt cgttgctgaa gcaggttggt ggcctagcgg
 38041 actcggagca gcccgaaact gaccccgccg aggtgcagct gcgcccgtcc gccctgtcgg
 38101 gggcagacca (SEQ ID NO: 24)

6.9. X-linked Inhibitor of Apoptosis Protein ("XIAP")

10 GenBank Accession # U45880:

1 gaaaagggtg acaagtccta tttcaagag aagatgactt ttaacagttt tgaaggatct
 61 aaaactgtg tacctgcaga catcaataag gaagaagaat ttgtagaaga gttaataga
 121 ttaaaaactt ttgctaattt tccaagtggg agtcctgttt cagcatcaac actggcacga
 15 181 gcaggggttc ttatactgg tgaaggagat accgtgcggt gctttagttg tcatgcagct
 241 gtagatagat ggcaatatgg agactcagca gttggaagac acaggaaagt atccccaagt
 301 tgcagattta tcaacggctt ttatctgaa aatagtcca cgcagtctac aaattctggt
 361 atccagaatg gtcagtacaa agttgaaaac tatctgggaa gcagagatca tttgcctta
 421 gacaggccat ctgagacaca tgcagactat cttttgagaa ctgggcaggt ttagatata
 20 481 tcagacacca tatacccgag gaaccctgcc atgtattgtg aagaagctag attaaagtcc
 541 ttcagaact ggccagacta tgctcaccta accccaagag agttagcaag tgctggactc
 601 tactacacag gtattggtga ccaagtgcag tgctttgtt gtggtggaaa actgaaaaa
 661 tgggaacctt gtatcgtgc ctggtcagaa cacaggcgac acttcctaa ttgctcttt
 721 gttttgggcc ggaatcttaa tattcgaagt gaatctgatg ctgtgagttc ttagaggaat
 25 781 ttcccaatt caacaaatc tcaagaaat ccatccatgg cagattatga agcacggatc
 841 ttacttttg ggacatggat atactcagtt aacaaggagc agcttgcaag agctggattt
 901 tatgctttag gtgaaggatg taaagtaaag tgctttcact gtggaggagg gctaactgat
 961 tggaagccca gtgaagacct ttgggaacaa catgctaaat ggtatccagg gtgcaaatat
 1021 ctgttagaac agaagggaca agaatatata aacaatattc atttaactca ttacttgag
 30 1081 gagtgtctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc
 1141 atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt
 1201 aagaaaataa tggaggaaaa aattcagata tctgggagca actataaatc acttgaggtt
 1261 ctggttgag atctagttaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact
 1321 tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt
 35 1381 tgcaaaatct gtatggatag aaatattgct atcgttttg ttcttggtg acatctagtc
 1441 acttgtaaac aatgtgctga agcagttgac aagtgtcca tgtgtacac agtcattact

1501 ttcaagcaaa aaatttttat gtcttaatct aactctatag taggcatgtt atgttgttct
 1561 tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat
 1621 tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata
 5 1681 atctttgaat ttcttgattt tcagggtat tagctgtatt atccattttt ttactgtta
 1741 ttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt
 1801 attcatagta tactgattta atttctaagt gtaagtgaat taatcatctg gattttttat
 1861 tcttttcaga taggcttaac aaatggagct ttctgtatat aaatgtggag attagagtta
 1921 atctcccaa tcacataaatt tgttttgtgt gaaaaaggaa taaattgttc catgctgggtg
 10 1981 gaaagataga gattgttttt agagggttgg tgttgtgttt taggattctg tccattttct
 2041 tgtaaaggga taaacacgga cgtgtgcgaa atatgtttgt aaagtgttt gccattgttg
 2101 aaagcgtatt taatgataga atactatcga gccacatgt actgacatgg aaagatgtca
 2161 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatagtctg agccagatca
 2221 aagtatgtat gttgttaata tgcatagaac gagagatttg gaaagatata caccaaactg
 15 2281 ttaaatgtgg ttctcttcg gggagggggg gatggggga ggggccccag aggggtttta
 2341 gaggggcctt ttactttcg actttttca tttgttctg ttggatttt ttataagtat
 2401 gtagaccccg aagggtttta tgggaactaa catcagtaac ctaaccccg tgactatcct
 2461 gtgctcttcc tagggagctg tgtgtttcc caccaccac ccttccctct gaacaaatgc
 2521 ctgagtgcgt gggcactttg (SEQ ID NO: 25)

20

General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:

5'AGCUCCUAUAACAAAAGUCUGUUGCUUGUGUUUCACAUUUUGGAUU
 UCCUAAUAUAAUGUUCUCUUUUUAGAAAAGGUGGACAAGUCCUAUUU
 25 UCAAGAGAAG3' (SEQ ID NO: 26)

25

Initial Specific Target Motif:

RNP core binding site within XIAP IRES

5'GGAUUUCCUAAUAUAAUGUUCUCUUUUU3' (SEQ ID NO: 27)

30

6.10. Survivin

GenBank Accession # NM_001168:

1 ccgccagatt tgaatcgagg gaccgttgg cagagggtgg ggcggcggca tgggtgcccc
 61 gacgttgccc cctgcctggc agccctttct caaggaccac cgcatctcta cattcaagaa
 121 ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctggttcat
 35 181 ccactgcccc actgagaacg agccagactt ggcccagtgt ttcttctgct tcaaggagct

241 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggtg
 301 cgctttcctt tctgtcaaga agcagtttga agaattaacc ctgtgtgaat tttgaaact
 361 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt
 5 421 tgaggaaact gcgaagaaag tgcgccgtgc catcgagcag ctggctgcca tggattgagg
 481 cctctggccg gagctgcctg gtcccagagt ggctgcacca ctccagggt ttattccctg
 541 gtgccaccag ccttcctgtg ggccccttag caatgtctta ggaaaggaga tcaacatttt
 601 caaattagat gttcaactg tgctcctgtt ttgtctttaa agtggcacca gaggtgcttc
 661 tgcctgtgca gcgggtgctg ctggtaacag tggctgcttc tctctctc tctctttt
 721 gggggctcat tttgctgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag
 10 781 aaggcagtgt ccttttgc agagctgaca gctttgttcg cgtgggcaga gcctccaca
 841 gtgaatgtgt ctggacctca tgtgttgag gctgtcacag tctgagtgt ggacttgga
 901 ggtgcctgtt gaatcigagc tgcaggtcc ttatctgca cacctgtgcc tctcagagg
 961 acagttttt tgtgtgtg ttttttgt tttttttt ggtagatgca tgacttgtt
 1021 gtgatgagag aatggagaca ggtccctgg ctctctact gtttaacaac atggctttt
 15 1081 tattttgtt gaattgttaa ttcacagaat agcacaaaact acaattaaaa ctaagcaca
 1141 agccattcta agtcattggg gaaacgggt gaacttcagg tggatgagga gacagaatag
 1201 agtgatagga agcgtctggc agatactcct ttgccactg ctgtgtgatt agacaggccc
 1261 agtgagccgc ggggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc
 1321 ctttttaaat gacttggtc gatgctgtgg gggactggct gggtgctgc aggccgtgtg
 20 1381 tctgtcagcc caaccttcac atctgtcacg ttctccacac gggggagaga cgagtcgcg
 1441 ccagggtccc gctttcttg gaggcagcag ctccgcagg gctgaagtct ggcgtaagat
 1501 gatggattg attcgccctc ctccctgtca tagagctgca ggtggattg ttacagcttc
 1561 gctggaaacc tctggaggtc atctcggtg ttctgagaa ataaaaagcc tgcatttc (SEQ ID NO: 28)

25 **7. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA**
BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS

The results presented in this Example indicate that gel mobility shift assays
 can be used to detect the binding of small molecules, such as the Tat peptide and
 30 gentamicin, to their respective target RNAs.

7.1. Materials and Methods

7.1.1. Buffers

35 Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH
 7.4, 20mM KCl, 0.1% Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is

composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1% Triton X-100 and 5mM MgCl₂.

5

7.1.1. Gel retardation analysis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of either a 5' fluorescein labeled oligonucleotide corresponding to the 16S rRNA A site (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 29); Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448) or a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was 3' labeled with 5'-³²P cytidine 3', 5'-bis(phosphate) (NEN) and T4 RNA ligase (NEBiolabs) in 10% DMSO as per manufacturer's instructions. The labeled oligonucleotides were purified using G-25 Sephadex columns (Boehringer Mannheim). For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). For 16S rRNA-gentamicin reactions, the method of Huq *et al.* was used with TKM buffer. In 20 µl reaction volumes 50 pmoles of ³²P cytidine-labeled oligonucleotide and either gentamicin sulfate (Sigma) or the short Tat peptide (Tat₄₇₋₅₈) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. Then 10 µl of 30% glycerol was added to each reaction tube and the entire sample was loaded onto a TBE non-denaturing polyacrylamide gel and electrophoresed at 1200-1600 volt-hours at 4°C. The gel was exposed to an intensifying screen and radioactivity was quantitated using a Typhoon phosphorimager (Molecular Dynamics).

30

7.2. Background

One method used to demonstrate small molecule interactions with natural occurring RNA structures such as ribosomes is by a method called chemical footprinting or toe printing (Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448). Here the use of gel mobility shift assays to monitor RNA-small molecule interactions are described. This approach allows for rapid visualization of small molecule-RNA interactions based on the

35

difference between mobility of RNA alone versus RNA in a complex with a small molecule. To validate this approach, an RNA oligonucleotide corresponding to the well-characterized gentamicin binding site on the 16S rRNA (Moazed & Noller, 1987, Nature 327:389-394) and the equally well-characterized HIV-1 TAT protein binding site on the HIV-1 TAR element (Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093) were chosen. The purpose of these experiments is to lay the groundwork for the use of chromatographic techniques in a high throughput fashion, such as microcapillary electrophoresis, for drug discovery.

7.3. Results

A gel retardation assay was performed using the Tat₄₇₋₅₈ peptide and the TAR RNA oligonucleotide. As shown in Figure 1, in the presence of the Tat peptide, a clear shift is visible when the products are separated on a 12% non-denaturing polyacrylamide gel. In the reaction that lacks peptide, only the free RNA is visible. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

Based on the results of Figure 1, it was hypothesized that RNA interactions with small organic molecules could also be visualized using this method. As shown in Figure 2, the addition of varying concentrations of gentamicin to an RNA oligonucleotide corresponding to the 16S rRNA A site produces a mobility shift. These results demonstrate that the binding of the small molecule gentamicin to an RNA oligonucleotide having a defined structure in solution can be monitored using this approach. In addition, as shown in Figure 2, a concentration as low as 10ng/ml gentamicin produces the mobility shift.

To determine whether lower concentrations of gentamicin would be sufficient to produce a gel shift, similar experiment was performed, as shown in Figure 2, except that the concentrations of gentamicin ranged from 100 ng/ml to 10 pg/ml. As shown in Figure 3, gel mobility shifts are produced when the gentamicin concentration is as low as 10 pg/ml. Further, the results shown in Figure 3 demonstrate that the shift is specific to the 16S rRNA oligonucleotide as the use of an unrelated oligonucleotide, corresponding to the HIV TAR RNA element, does not result in a gel mobility shift when incubated with 10 µg/ml gentamicin. In addition, if a concentration as low as 10 pg/ml gentamicin produces a gel mobility shift then it should be possible to detect changes to RNA structural motifs when small amounts of compound from a library of diverse compounds is screened in this fashion.

Further analysis of the gentamicin-RNA interaction indicates that the interaction is Mg²⁺- and temperature dependent. As shown in Figure 4, when MgCl₂ is not present (TK buffer), 1mg/ml of gentamicin must be added to the reaction to produce a gel shift.

Similarly, the temperature of the reaction when gentamicin is added is also important. When gentamicin is present in the reaction during the entire denaturation/renaturation cycle, that is, when gentamicin is added at 90°C or 85°C, a gel shift is visualized (data not shown). In contrast, when gentamicin is added after the renaturation step has proceeded to 75°C, a mobility shift is not produced. These results are consistent with the notion that gentamicin may recognize and interact with an RNA structure formed early in the renaturation process.

8. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS BY CAPILLARY ELECTROPHORESIS

The results presented in this Example indicate that interactions between a peptide and its target RNA, such as the Tat peptide and TAR RNA, can be monitored by gel retardation assays in an automated capillary electrophoresis system.

8.1. Materials and Methods

8.1.1. Buffers

Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl₂.

8.1.1. Gel Retardation Analysis Using Capillary Electrophoresis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was used. For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was

utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). In 20 µl reaction volumes 50 pmoles of labeled oligonucleotide and the short Tat peptide (Tat₄₇₋₅₈) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania).

8.2. Results

As presented in the previous Example in Section 7, interactions between a peptide and RNA can be monitored by gel retardation assays. It was hypothesized that interactions between a peptide and RNA could be monitored by gel retardation assays by an automated capillary electrophoresis system. To test this hypothesis, a gel retardation assay by an automated capillary electrophoresis system was performed using the Tat₄₇₋₅₈ peptide and the TAR RNA oligonucleotide. As shown in Figure 5 using the capillary electrophoresis system, in the presence of the Tat peptide, a clear shift is visible upon the addition of increasing concentrations of Tat peptide. In the reaction that lacks peptide, only a peak corresponding to the free RNA is observed. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entirety.

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed; (b) separating the detectably labeled target
10 RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds; and (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex.

 2. The method of paragraph 1 in which the target RNA molecule
15 contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.

 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6
20 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.

25 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.

 5. The method of paragraph 1 in which the test compound is selected
30 from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoin, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries, including but not limited to, libraries of benzodiazepines, isoprenoids, thiazolidinones,
35 metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

- 5 6. The method of paragraph 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions.
7. The method of paragraph 6 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.
- 10 8. The method of paragraph 6 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 15 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.
9. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. In a preferred embodiment, electrophoresis is used for separation of the complexed and non-complexed target nucleic acids. In a preferred embodiment, the electrophoresis is capillary electrophoresis. In other embodiments, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation are used for the separation of the complexed and non-complexed target nucleic acids.
- 25 10. The structure of the test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds.

35

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
- 5 (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed;
 - 10 (b) separating the detectably labeled target RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds by capillary gel electrophoresis; and
 - 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex by mass spectroscopy.
- 20
- 25
- 30
- 35

AMENDED CLAIMS

[received by the International Bureau on 17 September 2002 (17.09.02);
Claims 1 to 10 replaced by new claims 1 to 19. (3 sheets)]

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
- (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library
- 10 of test compounds so that a detectably labeled target RNA:test compound complex is formed;
- (b) separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and
- 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex.
2. The method of claim 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or
- 20 adenylate uridylate-rich element.
3. The method of claim 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6
- 25 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
4. The method of claim 1 in which the detectably labeled RNA is
- 30 labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
5. The method of claim 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as
- 35 hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal

peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; or small organic molecule libraries.

5

6. The method of claim 5 in which the small organic molecule libraries are libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

10

7. The method of claim 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution wherein the aqueous solution comprises a buffer and a combination of salts.

15

8. The method of claim 7 wherein the aqueous solution approximates or mimics physiologic conditions.

9. The method of claim 7 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

20

10. The method of claim 7 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant.

25

11. The method of claim 10 in which the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂.

13 The method of claim 11 wherein the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂.

30

14. The method of claim 10 wherein the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

35

15. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and
5 test compounds is by electrophoresis.
16. The method of claim 15 in which the electrophoresis is capillary electrophoresis.
- 10 17. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and test compounds is by fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, or
15 nanoparticle aggregation.
18. The method of claim 1 in which the library of test compounds are small organic molecule libraries.
- 20 19. The method of claim 18 in which the structure of the test compound is determined by mass spectroscopy, NMR, or vibration spectroscopy.

Figure 1

Sheet 1/5

Attorney Docket No. 10589-007

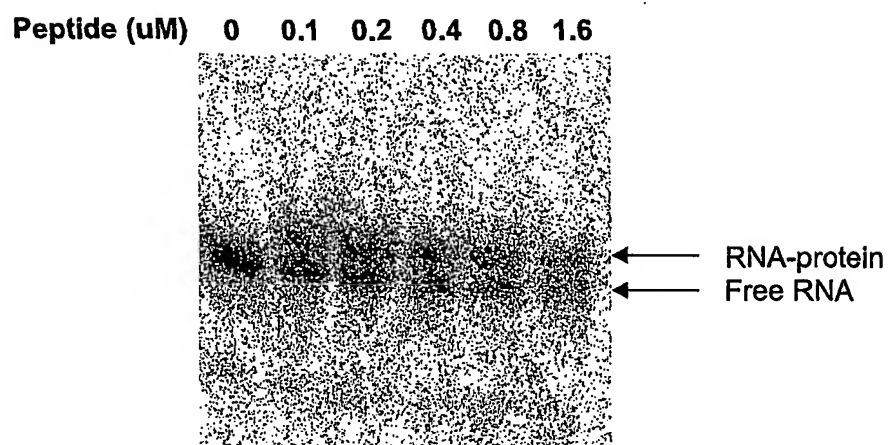


Figure 2

Sheet 2/5

Attorney Docket No. 10589-007

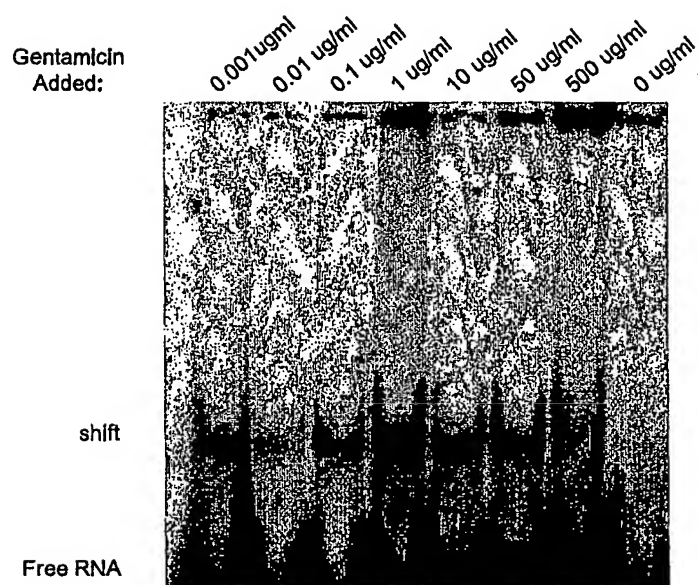


Figure 3

Sheet 3/5

Attorney Docket No. 10589-007

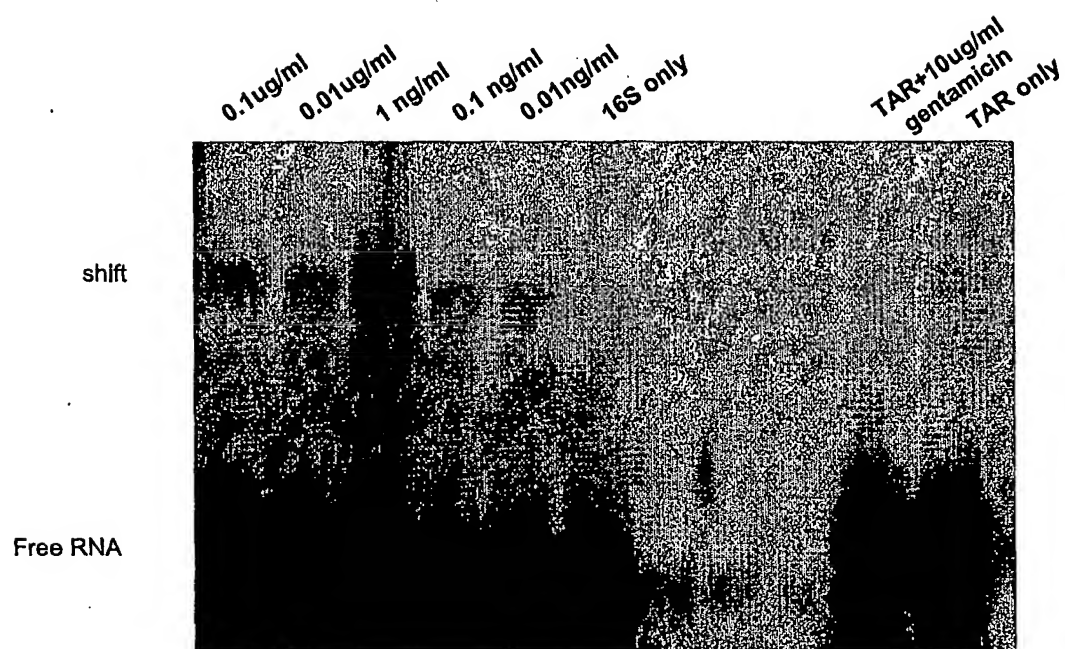


Figure 4

Sheet 4/5

Attorney Docket No. 10589-007

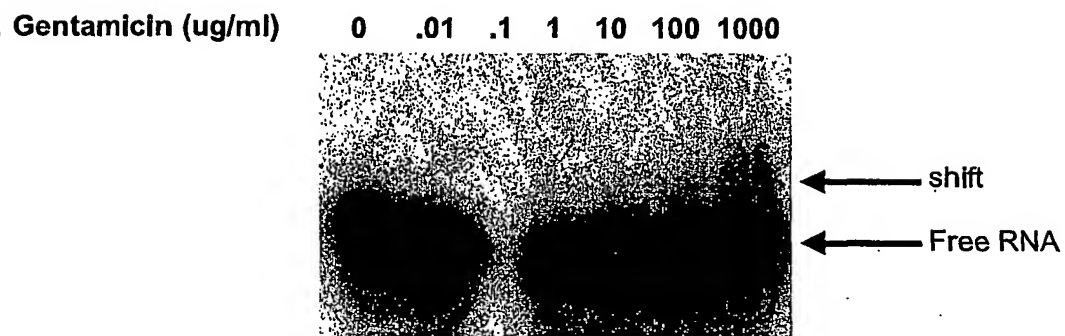
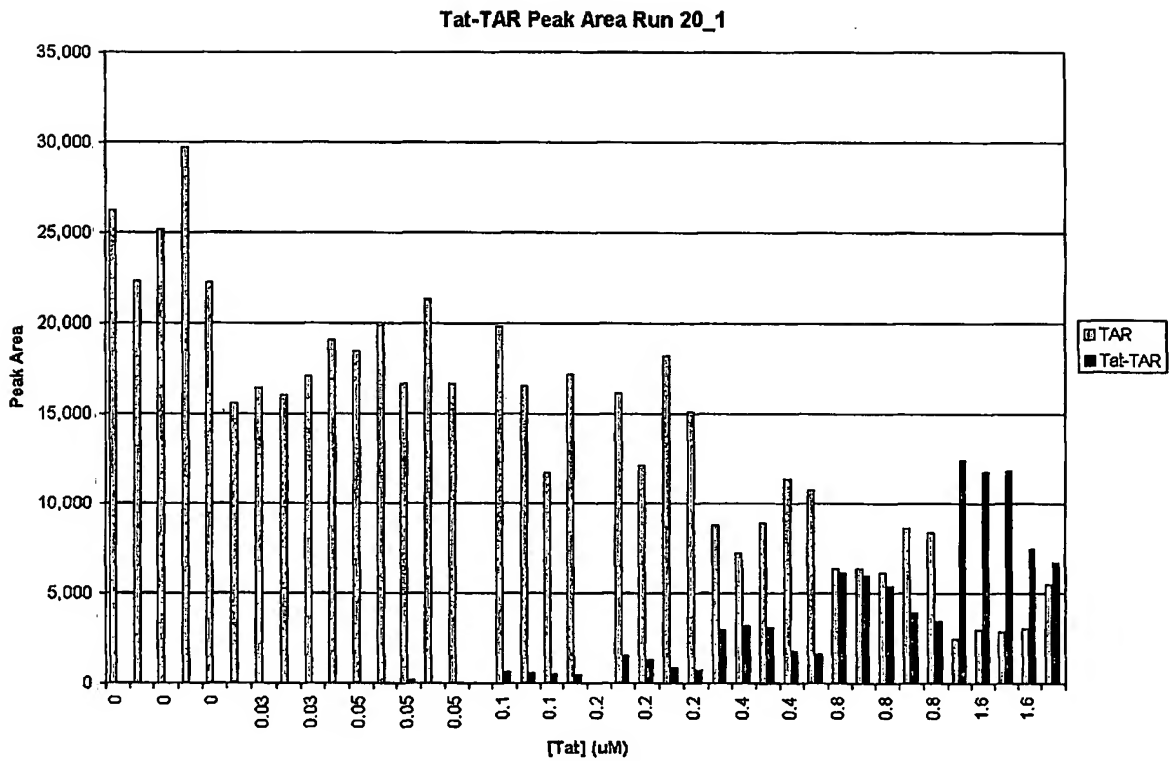


Figure 5

Sheet 5/5

Attorney Docket No. 10589-007



SEQUENCE LISTING

<110> PTC Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA
STRUCTURAL MOTIFS

<130> 10589-007-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,965

<151> 2001-04-11

<160> 31

<170> PatentIn version 3.0

<210> 1

<211> 21

<212> RNA

<213> Homo sapiens

<400> 1
auuuuuuuuu uuauuuuuuu a

21

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

<400> 2
auuuuuuuuu uuauuuua

17

<210> 3

<211> 15

<212> RNA

<213> Homo sapiens

<400> 3

wauuuuuuuu uuuw

15

<210> 4

<211> 13

<212> RNA

<213> Homo sapiens

<400> 4

wwauuuuuu aww

13

<210> 5

<211> 13

<212> RNA

<213> Homo sapiens

<400> 5

wwwauuuuaw www

13

<210> 6

<211> 1643

<212> DNA

<213> Homo sapiens

<400> 6

gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca 60

gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgaccac 120

ggctccaccc tctctcccct ggaaaggaca ccatgagcac tgaaagcatg atccgggacg 180

tggagctggc cgaggaggcg ctccccaaga agacaggggg gccccagggc tccaggcggt 240

gcttgttct cagcctcttc tccttctga tcgtggcagg cgccaccacg ctcttctgcc 300

tgctgcactt tggagtgatc ggccccaga ggaagagtt cccagggac ctctctctaa	360
tcagccctct ggcccaggca gtcagatcat cttctcgaac cccagtgac aagcctgtag	420
cccatgttgt agcaaaccct caagctgagg ggcagctcca gtggctgaac cgccgggcca	480
atgccctoct ggccaatggc gtggagctga gagataacca gctggtggtg ccatcagagg	540
gcctgtacct catctactcc caggctctct tcaagggcca aggetgccc tccacccatg	600
tgctcctcac ccacaccatc agccgcatcg ccgtctccta ccagaccaag gtcaacctcc	660
tctctgccat caagagcccc tgccagagg agaccccaga gggggtgag gccaaagccct	720
ggtatgagcc catctatctg ggaggggtct tcagctgga gaagggtgac cgactcagcg	780
ctgagatcaa tcggcccagc tatctcgact ttgccgagtc tgggcaggtc tactttggga	840
tcattgccct gtgaggagga cgaacatcca accttccaa acgctcccc tgcccaatc	900
cctttattac cccctccttc agacaccctc aacctcttct ggctcaaaaa gagaattggg	960
ggcttagggc cggaacccaa gcttagaact ttaagcaaca agaccaccac ttcgaaacct	1020
gggattcagg aatgtgtggc ctgcacagt aattgctggc aaccactaag aattcaaact	1080
ggggcctcca gaactcactg gggcctacag ctttgatccc tgacatctgg aatctggaga	1140
ccagggagcc tttggttctg gccagaatgc tgcaggactt gagaagacct cacctagaaa	1200
ttgacacaag tggaccttag gccttcctct ctccagatgt ttccagactt ccttgagaca	1260
cggagcccag ccctcccat ggagccagct cctctatatt atgtttgcac ttgtgattat	1320
ttattattta tttattattt atttatttac agatgaatgt atttatttgg gagaccgggg	1380
tatcctgggg gacccaatgt aggagctgcc ttggctcaga catgttttcc gtgaaaacgg	1440
agctgaacaa taggctgttc ccatgtagcc ccctggcctc tgtgccttct tttgattatg	1500
ttttttaaaa tatttatctg attaagttgt ctaacaatg ctgatttggg gaccaactgt	1560
cactcattgc tgagcctctg ctcccaggg gagttgtgtc tgtaatcgcc ctactattca	1620
gtggcgagaa ataaagtttg ctt	1643

<210> 7

<211> 756

<212> DNA

<213> Homo sapiens

<400> 7

gctggaggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg	60
cacccgcccg ctgcccagc cccagcacgc agccctggga gcatgtgaat gccatccagg	120
aggccggcg tctcctgaac ctgagtagag aactgctgc tgagatgaat gaaacagtag	180

aagtcacatc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc 240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg 300
ccagccacta caagcagcac tgccctccaa ccccggaac ttctgtgca acccagacta 360
tcacctttga aagtttcaaa gagaacctga aggactttct gcttgatc cctttgact 420
gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
tctctcatga aacaagagct agaaactcag gatggtcatc ttggaggag caaggggtgg 540
gccacagcca tgggtgggagt ggcctggacc tgccctgggc cacactgacc ctgatacagg 600
catggcagaa gaatgggaat atttatact gacagaaatc agtaatatat atatattat 660
atttttaaaa tatttattha tttatttatt taagttcata ttccatattt attcaagatg 720
ttttaccgta ataattatta ttaaaaatat gcttct , 756

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

<400> 8
tctggaggat gtggctgcag agcctgctgc tcttgggcac tgtggcctgc agcatctctg 60
caccgcccgc ctgcgccagc ccagcacgc agccctggga gcatgtgaat gccatccagg 120
aggccggcgg tctcctgaac ctgagtagag aactgctgc tgagatgaat gaaacagtag 180
aagtcacatc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc 240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg 300
ccagccacta caagcagcac tgccctccaa ccccggaac ttctgtgca acccagacta 360
tcacctttga aagtttcaaa gagaacctga aggactttct gcttgatc cctttgact 420
gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
tctctcatga aacaagagct agaaactcag gatggtcatc ttggaggag caaggggtgg 540
gccacagcca tgggtgggagt ggcctggacc tgccctgggc cacactgacc ctgatacagg 600
catggcagaa gaatgggaat atttatact gacagaaatc agtaatatat atatattat 660
atttttaaaa tatttattha tttatttatt taagttcata ttccatattt attcaagatg 720
ttttaccgta ataattatta ttaaaaatat gcttct 756

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

```

<400> 9
atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc      60
aactcctgtc ttgcattgca ctaattcttg cacttgtcac aaacagtgc cctacttcaa      120
gttcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggatttac      180
agatgatttt gaatggaatt aataattaca agaatcccaa actcaccagg atgctcacat      240
ttaagtttta catgcccaag aaggccacag aactgaaaca gcttcagtgt ctagaagaag      300
aactcaaacc tctggaggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac      360
ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa      420
cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga      480
ttacottttg tcaaagcatc atctcaacac taacttgata attaagtgt tccacttaa      540
aacatatcag gccttctatt tatttattta aatattttaa ttttatattt attgttgaat      600
gtatggttgc tacctattgt aactattatt cttaatctta aaactataaa tatggatctt      660
ttatgattct ttttgaagc cctaggggct ctaaaatggt ttaccttatt tatcccaaaa      720
atatttatta ttatgttgaa tgttaaatat agtatctatg tagatttggt agtaaaacta      780
ttaaataaat ttgataaata taaaaaaaaa aaacaaaaaa aaaaaa      825

```

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10
 nauuuuuuuu uuan

15

<210> 11

<211> 1125

<212> DNA

<213> Homo sapiens

<400> 11
 ttctgccttc gagcccaccg ggaacgaaag agaagctcta tctgcctcc aggagcccag 60
 ctatgaactc cttctccaca agcgccttcg gtccagttgc cttctccctg gggctgctcc 120
 tgggtgttgc tgctgccttc cctgccccag tacccccagg agaagattcc aaagatgtag 180
 ccgccccaca cagacagcca ctcacctctt cagaacgaat tgacaaaca attcgggtaca 240
 tctcgcacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa 300
 gcagcaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaagatg 360
 gatgcttcca atctggattc aatgaggaga cttgcctggt gaaaatcatc actggtcttt 420
 tggagtttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag 480
 ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga 540
 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc 600
 aggcacagaa ccagtggctg caggacatga caactcatct cattctgctc agctttaagg 660
 agttcctgca gtccagcctg agggctcttc ggcaaagtga gcatgggcac ctcagattgt 720
 tgttgttaat gggcattcct tcttctggtc agaaacctgt ccaactgggca cagaacttat 780
 gttgttctct atggagaact aaaagtatga gcgttaggac actattttta ttatttttaa 840
 tttattaata tttaaatatg tgaagctgag ttaatttatg taagtcatat ttatattttt 900
 aagaagtacc acttgaaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta 960
 tgcagtttga atatcctttg tttcagagcc agatcatttc ttggaaagtg taggcttacc 1020
 tcaaataaat ggctaactta tacatatatt taaagaaata tttatattgt atttatataa 1080
 tgtataaatg gtttttatac caataaatgg cattttaaaa aattc 1125

<210> 12

<211> 3166

<212> DNA

<213> Homo sapiens

<400> 12
aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcgggcgt 60
gcgagcagcg aaagcgacag gggcaaagt agtgaoctgc ttttgggggt gaccgccgga 120
gcgcggcggtg agccctcccc cttgggatcc cgcagctgac cagtcgcgct gacggacaga 180
cagacagaca ccgccccag cccagttac cacctcctcc ccggccggcg gcggacagtg 240
gacgcggcgg cgagccgcgg gcagggggccg gagcccgccc ccggaggcgg ggtggagggg 300
gtcggagctc gcggcgctgc actgaaactt ttctgtccaa ttctgggctg ttctcgttc 360
ggaggagccg tgggtccgcg gggggaagcc gagccgagcg gagccgcgag aagtgctagc 420
tcgggccggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag 480
agggggccgc agtggcgact cggcgctcgg aagccgggct catggacggg tgaggcgcg 540
gtgtgcgag acagtgtctc agcgcgcgag ctcccagcc ctggcccggc ctccggccgg 600
gaggaagagt agctcgccga ggcgcgag agagcgggccc gcccacagc ccgagccgga 660
gagggacgag agccgcgcgc ccgggtcggg cctccgaaac catgaacttt ctgctgtctt 720
gggtgcattg gagccttgcc ttgctgtctt acctccacca tgccaagtgg tcccaggctg 780
caccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct 840
atcagcgcag ctactgcat ccaatcgaga cctgggtgga catcttcag gagtaccctg 900
atgagatoga gtacatcttc aagccatcct gtgtgcccct gatgcgatgc gggggctgct 960
ccaatgacga gggcctggag tgtgtgcccc ctgaggagtc caacatcacc atgcagatta 1020
tgcgatcaa acctcacaa ggcagcaca taggagagat gagcttccta cagcacaaca 1080
aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccttgct 1140
cagagcggag aaagcatttg tttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa 1200
acacacactc gcgttgcaag gcgaggcagc ttgagttaaa cgaacgtact tgcagatgtg 1260
acaagccgag gcggtgagcc gggcaggag aaggagcctc cctcagggtt tcgggaacca 1320
gatctctctc caggaaagac tgatacagaa cgatcgatac agaaaccacg ctgccgccac 1380
cacaccatca ccatcgacag aacagtcctt aatccagaaa cctgaaatga aggaagagga 1440
gactctgcgc agagcacttt gggccggag ggcgagactc cggcgggaagc attccggggc 1500
gggtgaccca gcacggtccc tcttggaatt ggattcgcca ttttattttt cttgtgtgta 1560
aatcacgag ccggaagat tagagagttt tatttctggg attcctgtag acacaccac 1620
ccacatacat acatttatat atatatatata tatatatata taaaataaa tatctctatt 1680
ttatatatat aaaatatata tattcttttt ttaaattaac agtgctaatt ttattggtgt 1740
cttactgga tgtatttgac tgctgtggac ttgagttggg aggggaatgt tccactcag 1800

atcctgacag ggaagaggag gagatgagag actctggcat gatctttttt ttgtcccact 1860
 tgggtggggcc agggtcctct cccctgcca agaattgtga aggccagggc atgggggcaa 1920
 atatgacca gttttggaa caccgacaaa ccagccctg gcgctgagcc tctctacccc 1980
 aggtcagacg gacagaaaga caaatcacag gttccgggat gaggacaccg gctctgacca 2040
 ggagtgtggg gagcttcagg acattgctgt gctttggga ttccctccac atgctgcacg 2100
 cgcctctcgc cccaggggc actgcctgga agattcagga gcctgggagg ccttcgctta 2160
 ctctcacctg cttctgagtt gccagaggg ccactggcag atgtcccggc gaagagaaga 2220
 gacacattgt tggaagaagc agcccatgac agcgcctt cctgggactc gccctcatcc 2280
 tcttctgct ccccttcctg ggggtgcagcc taaaaggacc tatgtctca caccattgaa 2340
 accactagtt ctgtccccc aggaacctg gttgtgtgtg tgtgagtggg tgaccttct 2400
 ccatccctg gtccttcctc tccctcccg aggcacagag agacagggca ggatccacgt 2460
 gccattgtg gaggcagaga aaagagaaag tgttttatat acggtactta tttaatatcc 2520
 ctttttaatt agaaattaga acagttaatt taattaaaga gttagggtttt ttttcagtat 2580
 tcttggttaa tatttaattt caactattta tgagatgtat cttttgctct ctcttgctct 2640
 cttatttgta ccggtttttg tatataaaat tcatgtttcc aatctctctc tccctgatcg 2700
 gtgacagtca ctagcttatc ttgaacagat atttaatttt gctaacactc agctctgccc 2760
 tcccgatcc cctggctccc cagcacacat tcctttgaaa gagggtttca atatacatct 2820
 acatactata tatatatagg gcaacttgta tttgtgtgta tatatatata tatatgttta 2880
 tgtatatatg tgatcctgaa aaaataaaca tcgctattct gttttttata tgttcaaacc 2940
 aaacaagaaa aaatagagaa ttctacatac taaatctctc tcctttttta attttaatat 3000
 ttgttatcat ttattttattg gtgctactgt ttatccgtaa taattgtggg gaaaagatat 3060
 taacatcaag tctttgtctc tagtgcagtt ttctgagata ttccgtagta catatttatt 3120
 tttaaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaa 3166

<210> 13

<211> 249

<212> RNA

<213> Homo sapiens

<400> 13

ccgggcucau ggacggguga ggcggcgug ugcgagaca gugcuccagc gcgcgcgcuc 60
 ccagcccug gcccgccuc gggccgggag gaagaguagc ucgccgaggc gccgaggaga 120
 gcgggcccgc ccacagcccg agccggagag ggacgcgagc cgcgcgcccc ggucgggccu 180

ccgaaacc au gaacuuucug cugucuuggg ugcauuggag ccuugccuug cugcucuacc 240
uccaccaug 249

<210> 14

<211> 9181

<212> DNA

<213> Homo sapiens

<400> 14

ggtctctctg gttagaccag atctgagcct gggagctctc tggctaacta gggaaccac 60
tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt 120
gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca 180
gtggcgcccg aacagggacc tgaaagcgaa agggaaacca gaggagctct ctgcacgcag 240
gactcggcctt gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc 300
aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa 360
gcgggggaga attagatcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat 420
ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg 480
gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc 540
agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc 600
atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa 660
acaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggacacagc aatcagggtca 720
gccaaaatta ccctatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac 780
ctagaacttt aaatgcatgg gtaaaagtag tagaagagaa ggctttcagc ccagaagtga 840
tacctatggtt ttcagcatta tcagaaggag ccaccccaca agatttaaac accatgctaa 900
acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaag 960
ctgcagaatg ggatagagtg catccagtgc atgcagggcc tattgcacca ggccagatga 1020
gagaaccaag gggaagtgc atagcaggaa ctactagtag ccttcaggaa caaataggat 1080
ggatgacaaa taatccacct atcccagtag gagaaattta taaaagatgg ataatcctgg 1140
gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac 1200
caaaggaacc ctttagagac tatgtagacc ggttctataa aactctaaga gccgagcaag 1260
cttcacagga ggtaaaaaat tggatgacag aaacottggtt ggtccaaaat gcgaaccacg 1320
attgtaagac tatttttaaaa gcattgggac cagcggctac actagaagaa atgatgacag 1380
catgtcaggg agtaggagga cccggccata aggcaagagt tttggctgaa gcaatgagcc 1440

aagtaacaaa ttcagctacc ataatgatgc agagaggcaa ttttaggaac caaagaaaga	1500
ttgttaagtg tttcaattgt ggcaaagaag ggcacacagc cagaaattgc agggccccta	1560
ggaaaaaggg ctgttggaag tgtggaaagg aaggacacca aatgaaagat tgtactgaga	1620
gacaggctaa ttttttaggg aagatctggc cttcctacaa gggaaggcca gggaattttc	1680
ttcagagcag accagagcca acagccccac cagaagagag cttcaggtct ggggtagaga	1740
caacaactcc ccctcagaag caggagccga tagacaagga actgtatcct ttaacttccc	1800
tcaggtcact ctttggaac gacccctcgt cacaataaag ataggggggc aactaaagga	1860
agctctatta gatacaggag cagatgatac agtattagaa gaaatgagtt tgccaggaag	1920
atggaaacca aaaatgatag ggggaattgg aggttttatac aaagtaagac agtatgatca	1980
gatactcata gaaatctgtg gacataaagc tataggtaca gtattagtag gacctacacc	2040
tgtcaacata attggaagaa atctgttgac tcagattggg tgcactttaa attttcccat	2100
tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa	2160
acaatggcca ttgacagaag aaaaaataaa agcattagta gaaatttgta cagagatgga	2220
aaaggaaggg aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagtatttgc	2280
cataaagaaa aaagacagta ctaaattggag aaaattagta gatttcagag aacttaataa	2340
gagaactcaa gacttctggg aagttcaatt aggaatacca catcccgag ggtaaaaaa	2400
gaaaaaatca gtaacagtac tggatgtggg tgatgcata ttttcagttc ccttagatga	2460
agacttcagg aagtatactg catttaccat acctagtata aacaatgaga caccagggat	2520
tagatatcag tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag	2580
tagcatgaca aaaatcttag agccttttag aaaacaaaat ccagacatag ttatctatca	2640
atacatggat gatttgtatg taggatctga cttagaaata gggcagcata gaacaaaaat	2700
agaggagctg agacaacatc tggtgaggtg gggacttacc acaccagaca aaaaacatca	2760
gaaagaacct ccattccttt ggatgggtta tgaactccat cctgataaat ggacagtaca	2820
gcctatagtg ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg	2880
gaaattgaat tgggcaagtc agatttacc aggatataa gtaaggcaat tatgtaaact	2940
ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga	3000
actggcagaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgacctatc	3060
aaaagactta atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaattta	3120
tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaattgagg gtgcccacac	3180
taatgatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat	3240
atggggaaag actcctaaat ttaactgcc catacaaaag gaaacatggg aaacatgggtg	3300
gacagagtat tggcaagcca cctggattcc tgagtgggag tttgttaata cccctccctt	3360

agtgaaatta tggtaccagt tagagaaaga acccatagta ggagcagaaa ccttctatgt 3420
 agatggggca gctaacaggg agactaaatt aggaaaagca ggatatgtta ctaatagagg 3480
 aagacaaaaa gttgtcacc taactgacac aacaaatcag aagactgagt tacaagcaat 3540
 ttatctagct ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc 3600
 attaggaatc attcaagcac aaccagatca aagtgaatca gagttagtca atcaaataat 3660
 agagcagtta ataaaaaagg aaaaggtcta tctggcatgg gtaccagcac acaaaggaat 3720
 tggaggaaat gaacaagtag ataaattagt cagtgtctga atcaggaaag tactatTTTT 3780
 agatggaata gataaggccc aagatgaaca tgagaaatat cacagtaatt ggagagcaat 3840
 ggctagtgat tttaacctgc cacctgtagt agcaaaagaa atagtagcca gctgtgataa 3900
 atgtcagcta aaaggagaag ccatgcatgg acaagtagac tgtagtcag gaatatggca 3960
 actagattgt acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg 4020
 atatatagaa gcagaagtta ttccagcaga aacagggcag gaaacagcat attttctttt 4080
 aaaattagca ggaagatggc cagtaaaaac aatacatact gacaatggca gcaatttcac 4140
 cgggtgctacg gttagggccg cctgttggtg ggcggaatc aagcaggaat ttggaattcc 4200
 ctacaatccc caaagtcaag gagtagtaga atctatgaat aaagaattaa agaaaattat 4260
 aggacaggta agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat 4320
 ccacaatttt aaaagaaaag gggggatttg ggggtacagt gcaggggaaa gaatagtaga 4380
 cataatagca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa 4440
 ttttcgggtt tattacaggg acagcagaaa tccactttgg aaaggaccag caaagctcct 4500
 ctggaaaggt gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaag 4560
 aagaaaagca aagatcatta gggattatgg aaaacagatg gcagggtgatg attgtgtggc 4620
 aagtagacag gatgaggatt agaacatgga aaagtttagt aaaacaccat atgtatgttt 4680
 cagggaaaagc taggggatgg ttttatagac atcactatga aagccctcat ccaagaataa 4740
 gttcagaagt acacatccca ctaggggatg ctagattggg aataacaaca tattgggggc 4800
 tgcatacagg agaaagagac tggcatttgg gtcaggaggt ctccatagaa tggaggaaaa 4860
 agagatatag cacacaagta gaccctgaac tagcagacca actaattcat ctgtattact 4920
 ttgactgttt ttcagactct gctataagaa aggccttatt aggacacata gttagcccta 4980
 ggtgtgaata tcaagcagga cataacaagg taggatctct acaatacttg gcactagcag 5040
 cattaataac accaaaaaag ataaagccac ctttgctag tgttacgaaa ctgacagagg 5100
 atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggac 5160
 actagagctt ttagaggagc ttaagaatga agctgttaga cattttccta ggatttggct 5220
 ccatggctta gggcaacata tctatgaaac ttatggggat acttgggcag gagtgggaagc 5280

cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgtcgacata 5340
gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tcctagacta 5400
gagccctgga agcatccagg aagtcagcct aaaactgctt gtaccaattg ctattgtaaa 5460
aagtgttgct ttcattgcca agtttgtttc ataacaaaag ccttaggcat ctccatggc 5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct 5580
ctatcaaagc agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca 5640
ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg 5700
aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa 5760
gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagatg 5820
gggcaccatg ctccctggga tgttgatgat ctgtagtgtc acagaaaaat tgtgggtcac 5880
agtctattat ggggtacctg tgtggaagga agcaaccacc actctatttt gtgcatcaga 5940
tgctaaagca tatgatacag aggtacataa tgtttgggcc acacatgcct gtgtaccac 6000
agacccaac ccacaagaag tagtattggt aaatgtgaca gaaaatttta acatgtggaa 6060
aaatgacatg gtagaacaga tgcattgagga tataatcagt ttatgggatc aaagcctaaa 6120
gccatgtgta aaattaaccc cactctgtgt tagtttaaaag tgcaactgatt tgaagaatga 6180
tactaatacc aatagtagta gcgggagaat gataatggag aaaggagaga taaaaaactg 6240
ctctttcaat atcagcacia gcataagagg taagggtgag aaagaatatg cattttttta 6300
taaacttgat ataataccaa tagataatga tactaccagc tataagttga caagttgtaa 6360
cacctcagtc attacacagg cctgtccaaa ggtatccttt gagccaattc ccatacatta 6420
ttgtgccccg gctgggtttg cgattctaaa atgtaataat aagacgttca atggaacagg 6480
accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac 6540
tcaactgctg ttaaattgca gtctagcaga agaagaggta gtaattagat ctgtcaattt 6600
cacggacaat gctaaaacca taatagtaca gctgaacaca tctgtagaaa ttaattgtac 6660
aagacccaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt 6720
tgttacaata ggaaaaatag gaaatatgag acaagcacat tgtaacatta gtagagcaaa 6780
atggaataac actttaaaac agatagctag caaattaaga gaacaatttg gaaataataa 6840
aacaataatc ttttaagcaat cctcaggagg ggaccagaa attgtaacgc acagttttta 6900
ttgtggaggg gaatttttct actgtaattc aacacaactg tttaatagta cttggtttta 6960
tagtacttgg agtactgaag ggtcaaataa cactgaagga agtgacacaa tcaccctccc 7020
atgcagaata aaacaaatta taaacatgtg gcagaaagta ggaaaagcaa tgtatgcccc 7080
tcccatcagt ggacaaatta gatgttcac aaatattaca gggctgctat taacaagaga 7140
tggtggtaat agcaacaatg agtccgagat cttcagacct ggaggaggag atatgaggga 7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc	7260
accaccaag gcaaagagaa gagtgggtgca gagagaaaaa agagcagtgg gaataggagc	7320
tttgttcctt gggttcttgg gaggcagcagg aagcactatg ggcgcagcct caatgacgct	7380
gacggtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag	7440
ggctattgag gcgcaacagc atctgttgca actcacagtc tggggcatca agcagctcca	7500
ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg ggatttgggg	7560
ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgctagtt ggagtaataa	7620
atctctggaa cagatttggga atcacacgac ctggatggag tgggacagag aaattaacaa	7680
ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga	7740
acaagaatta ttggaattag ataaatgggc aagtttgtgg aatttggtta acataacaaa	7800
ttggctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gtttaagaat	7860
agtttttgct gtactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt	7920
tcagaccac ctccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg	7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg	8040
ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat	8100
tgtaacgagg attgtggaac ttctgggacg caggggggtg gaagccctca aatattggtg	8160
gaatctccta cagtattgga gtcaggaact aaagaatagt gctgttagct tgctcaatgc	8220
cacagccata gcagtagctg aggggacaga taggggtata gaagtagtac aaggagcttg	8280
tagagctatt cgccacatac ctagaagaat aagacagggc ttggaaagga ttttgctata	8340
agatgggtgg caagtggta aaaagtagtg tgattggatg gcctactgta agggaaagaa	8400
tgagacgagc tgagccagca gcagataggg tgggagcagc atctcgagac ctggaaaaac	8460
atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttggtgc tggctagaag	8520
cacaagagga ggaggagggt ggttttccag tcacacctca ggtacctta agaccaatga	8580
cttacaaggc agctgtagat cttagccact ttttaaaaga aaagggggga ctggaagggc	8640
taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct	8700
acttcctga ttagcagaac tacacaccag ggccaggggt cagatatcca ctgaccttg	8760
gatggtgcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag	8820
agaacaccag cttgttacac cctgtgagcc tgcattggat ggatgaccg gagagagaag	8880
tgtagagtg gaggtttgac agccgcctag catttcatca cgtggccga gagctgcac	8940
cggagtactt caagaactgc tgacatcgag cttgctacaa gggactttcc gctggggact	9000
ttccagggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tcctgcatat	9060
aagcagctgc tttttgcctg tactgggtct ctctgggttag accagatctg agcctgggag	9120

ctctctggct aactagggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgtt 9180
c 9181

<210> 15

<211> 29

<212> RNA

<213> Homo sapiens

<400> 15
ggcagaucug agccugggag cucucugcc 29

<210> 16

<211> 52

<212> RNA

<213> Homo sapiens

<400> 16
uuuuuuaggg aagaucuggc cuuccuacaa gggaaggcca gggaauuuuc uu 52

<210> 17

<211> 9413

<212> DNA

<213> Homo sapiens

<400> 17
ttgggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga 60
aagcgtctag ccatggcggt agtatgagt ttgtgcagcc tccaggaccc cccctcccg 120
gagagccata gtggtctgcg gaaccggtga gtacaccgga attgccagga cgaccgggtc 180
ctttcttggg tcaaccgct caatgcctgg agatttgggc gtgccccgc gagactgcta 240
gccgagtagt gttgggtcgc gaaaggcctt gtggtactgc ctgatagggt gcttgcgagt 300
gcccggggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaagaaaaa 360
ccaaacgtaa caccaaccgc cgcccacagg acgttaagtt cccgggcggt ggtcagatcg 420
ttggtggagt ttacctgttg ccgcgcaggg gcccaggtt ggggtgtgcgc gcgactagga 480
agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatcccaag gctcgccggc 540
ccgagggtag gacctgggct cagcccggtt acccttggcc cctctatggc aacgagggtg 600

tggggtgggc aggatggctc ctgtcacccc gtggtctctg gcctagttag ggccccacag	660
acccccggcg taggtcgcggt aatttgggta aggtcatcga tacccttaca tgcggcttcg	720
ccgacctcat ggggtacatt ccgcttgctg gcgccccct agggggcgct gccagggccc	780
tggcacatgg tgtccgggtt ctggaggacg gcgtgaacta tgcaacaggg aatctgcccg	840
gttgctcttt ctctatcttc ctottagctt tgcgtgtctt tttgaccatc ccagcttcg	900
cttacgaggt gcgcaacgtg tccgggatat accatgtcac gaacgactgc tccaactcaa	960
gtattgtgta tgaggcagcg gacatgatca tgcacacccc cgggtgcgtg ccctgcgtcc	1020
gggagagtaa tttctcccggt tgcgtggtag cgctcactcc cacgctcgcg gccaggaaca	1080
gcagcatccc caccacgaca atacgacgcc acgtcgattt gctcgttggg gcggctgctc	1140
tctgttccgc tatgtacgtt ggggatctct gcggatccgt tttctcgtc tcccagctgt	1200
tcaccttctc acctcgccgg tatgagacgg tacaagattg caattgctca atctatcccg	1260
gccacgtatc aggtcacgcg atggcttggg atatgatgat gaactgggtc cctacaacgg	1320
ccctagtgggt atgcagcta ctccggatcc cacaagccgt cgtggacatg gtggcggggg	1380
cccactgggg tgcctagcg ggccttgcc actattocat ggtggggaac tgggctaagg	1440
tcttgattgt gatgtactc tttgctggcg ttgacgggca caccacgtg acagggggaa	1500
gggtagcctc cagcaccag agcctcgtgt cctggctctc acaaggcca tctcagaaaa	1560
tccaactcgt gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg	1620
actccctcca aactgggttc attgctgcgc tgttctacgc acacagggtc aacgcgtccg	1680
ggtgcccaga gcgcatggct agctgccgcc ccacgatga gttcgctcag ggggtggggtc	1740
ccatcactca tgatatgcct gagagctcgg accagaggcc atattgctgg cactacgcgc	1800
ctcgaccgtg cgggatcgtg cctgcgtcgc aggtgtgtgg tccagtgtat tgcttcactc	1860
cgagccctgt tgtagtggg acgaccgatc gtttcggcgc tctacgtat agctgggggg	1920
agaatgagac agacgtgctg ctacttagca acacgcggcc gcctcaaggc aactggtttg	1980
ggtgcacgtg gatgaacagc actgggttca ccaagacgtg cgggggccct ccgtgcaaca	2040
tgggggggt cggcaacaac accttgggtc gccccacgga ttgcttcgg aagcaccocg	2100
aggccactta cacaagtgt ggctcggggc cctggttgac acccaggtgc atggttgact	2160
accdatacag gctctggcac taccctgca ctgttaactt taccgtcttt aaggtcagga	2220
tgtatgtggg gggcgtggag cacaggctca atgctgcatg caattggact cgaggagagc	2280
gctgtgactt ggaggacagg gataggtcag aactcagccc gctgctgctg tctacaacag	2340
agtggcagat actgcctgt tcttcacca ccctaccggc cctgtccact ggcttgatcc	2400
atcttcaccg gaacatcgtg gacgtgcaat acctgtacgg tatagggtcg gcagttgtct	2460
cctttgcaat caaatgggag tatatcctgt tgcttttcct tcttctggcg gacgcgcgcg	2520

tctgtgcctg	cttgtggatg	atgctgctga	tagcccaggc	tgaggccacc	ttagagaacc	2580
tggtggtcct	caatgcggcg	tctgtggccg	gagcgcatgg	ccttctctcc	ttcctcgtgt	2640
tcttctgctg	cgcctggtac	atcaaaggca	ggctggtccc	tggggcggca	tatgctctct	2700
atggcgatg	gcggttgctc	ctgctcttgc	tggccttacc	accacgagct	tatgccatgg	2760
accgagagat	ggctgcatcg	tgcggaggcg	cggtttttgt	aggtctggta	ctcttgacct	2820
tgtcaccata	ctataagggtg	ttcctcgcta	ggctcatatg	gtggttacaa	tattttatca	2880
ccagagccga	ggcgcacttg	caagtgtggg	tccccctct	caatgttcgg	ggaggccgcg	2940
atgccatcat	cctccttaca	tgcgcggtcc	atccagagct	aatccttgac	atcaccaaac	3000
tcctgctcgc	catactcggt	ccgctcatgg	tgctccaggc	tggcataact	agagtgcctg	3060
actttgtacg	cgctcagggg	ctcatccgtg	catgcatgtt	agtgcggaag	gtcgtcggag	3120
gccactatgt	ccaaatggcc	ttcatgaagc	tggccgcgct	gacaggtagc	tacgtatatg	3180
accatcttac	tccactgcgg	gattgggccc	acgcgggcct	acgagacctt	gcggtggcag	3240
tagagcccgt	cgtcttctct	gacatggaga	ctaaactcat	cacctggggg	gcagacaccg	3300
cggcgtgtgg	ggacatcatc	tcgggtctac	cagtctccgc	ccgaaggggg	aaggagatac	3360
ttctaggacc	ggccgatagt	tttggagagc	aggggtggcg	gctccttgcg	cctatcacgg	3420
cctattccca	acaaacgcgg	ggcctgcttg	gctgtatcat	cactagcctc	acaggtcggg	3480
acaagaacca	ggtcgatggg	gaggttcagg	tgctctccac	cgcaacgcaa	tctttcctgg	3540
cgacctgcgt	caatggcggtg	tgttggaacc	tctaccatgg	tgcgggctcg	aagaccctgg	3600
ccggcccga	gggtccaatc	acccaaatgt	acaccaatgt	agaccaggac	ctcgtcggct	3660
ggccggcgcc	ccccggggcg	cgctccatga	caccgtgcac	ctgcggcagc	tcggaccttt	3720
acttggtcac	gaggcatgct	gatgtcgttc	cgggtgcgcg	gcggggcgac	agcaggggga	3780
gcctgctttc	ccccaggccc	atctcctacc	tgaagggctc	ctcgggtgga	ccactgcttt	3840
gcccttcggg	gcacgttgta	ggcatcttcc	gggctgctgt	gtgcaccgga	ggggttcgga	3900
aggcggtgga	cttcataccc	gttgagtcta	tggaaactac	catgcgggtc	ccggtcttca	3960
cagacaactc	atcccctccg	gccgtaccgc	aaacattcca	agtggcacat	ttacacgctc	4020
ccactggcag	cggcaagagc	accaaagtgc	cggctgcata	tgcagcccaa	gggtacaagg	4080
tgctcgtcct	aaaccgcgtc	gttgccgcca	cattgggctt	tggagcgtat	atgtccaagg	4140
cacatggcat	cgagcctaac	atcagaactg	gggtaaggac	catcaccacg	ggcggcccca	4200
tcaogtactc	cacctattgc	aagttccttg	ccgacggtgg	atgctccggg	ggcgcctatg	4260
acatcataat	atgtgatgaa	tgccactcaa	ctgactcgac	taccatcttg	ggcatcggca	4320
cagtccctgga	tcaggcagag	acggctggag	cgcggctcgt	cgtgctcgcc	accgccacgc	4380
ctccgggata	gatcaccgtg	ccacacccca	acatcgagga	agtggccctg	tccaacactg	4440

gagagattcc cttctatggc aaagccatcc ccattgaggc catcaagggg ggaaggcatc 4500
tcattcttctg ccattccaag aagaagtgtg acgagctcgc cgcaaagctg acaggcctcg 4560
gactcaatgc tgtagcgtat taccggggtc tcgatgtgtc cgtcataccg actagcggag 4620
acgtcgttgt cgtggcaaca gacgctctaa tgacgggttt taccggcgac tttgactcag 4680
tgatcgactg caacacatgt gtcacccaga cagtcgattt cagcttgat cccaccttca 4740
ccattgagac gacaacgctg cccaagacg cgggtgtcgcg tgcgcagcgg cgaggtagga 4800
ctggcagggg caggagtggc atctacaggt ttgtgactcc aggagaacgg ccctcaggca 4860
tgttcgactc ctoggtcctg tgtgagtgt atgacgcagg ctgcgcttgg tatgagctca 4920
cgcccgctga gacctcggtt aggttgccgg cttacctaaa tacaccaggg ttgcccgctt 4980
gccaggacca cctagagttc tgggagagcg tcttcacagg cctcaccac atagatgcc 5040
acttcttctc ccagacaaa caggcaggag acaacctccc ctacctggtg gcataccaag 5100
ccacagtgtg cgccagggtc caggctccac ctccatcgtg ggaccaaag tggaagtgtc 5160
tcatacggct aaagcccaca ctgcatgggc caacggccct gctgtacagg ctaggagccg 5220
ttcaaatga ggtcactctc acacacccca taaccaaata catcatggca tgcattgtcg 5280
ctgacctgga ggtcgtcact agcacctggg tgctagtagg cggagtcctt gcggctctgg 5340
ccgcttactg cctgacgaca ggcagcgtg tcatgtgtgg caggatcatc ttgtccggga 5400
ggccagctgt tattcccgac agggaagtcc tctaccagga gttcgatgag atggaagagt 5460
gtgcttcaca cctcccttac atcgagcaag gaatgcagct cgccgagcaa ttcaaacaga 5520
aggcgtcgg attgctgcaa acagccacca agcaagcggg ggctgctgct ccggtgggtg 5580
agtccaagtg gcgagccctt gaggtcttct gggcgaaaca catgtggaac ttcatcagcg 5640
ggatacagta cttggcaggc ctatccactc tgcttgaaa ccccgcgata gcatcattga 5700
tggtttttac agcctctatc accagcccgc tcaccacca aaataccctc ctgtttaaca 5760
tcttgggggg atgggtggct gcccaactcg ctccccccag cgctgcttcg gctttcgtgg 5820
gcgcggcat tgccggtgcg gccgttgga gcataggtct cgggaaggta cttgtggaca 5880
ttctggcggg ctatggggcg ggggtggctg gcgcactcgt ggctttaaag gtcattgagc 5940
gcgagatgcc ctccactgag gatctggtta atttactccc tgccatcctt tctcctggcg 6000
ccctgggtgt cgggtcgtg tgcgcagcaa tactgcgtcg gcaogtgggc ccgggagagg 6060
gggctgtgca gtggatgaac cggctgatag cgttcgcttc gcggggtaac cacgtctccc 6120
ccacgcacta tgtgcccgag agcgacgccg cggcgcgtgt tactcagatc ctctccagcc 6180
ttaccatcac tcagttgctg aagaggttc atcagtgat taatgaggac tgctccaogc 6240
cttgttcgg ctctgggcta aaggatgttt gggactggat atgcacgggtg ttgagtgact 6300
tcaagacttg gctccagtcc aagctcctgc cgcggttacc ggaactccct ttcctgtcat 6360

gccaacgogg gtacaaggga gtctggcggg gggatggcat catgcaaacc acctgcccac 6420
 gtggagcaca gatcacggga catgtcaaaa atggctccat gaggattgtt gggccaaaaa 6480
 cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc atacaccacg ggcccctgca 6540
 cgccctcccc agcgccgaac tattccaggg cgctgtggcg ggtggctgct gaggagtacg 6600
 tggaggttac gggggtgggg gatttccact acgtgacggg catgaccact gacaacgtga 6660
 aatgcccacg ccaggttcca gcccctgaat ttttcacgga ggtggatgga gtacggttgc 6720
 acaggtatgc tccagtgtgc aaacctctcc tacgagagga ggtcgtattc caggtcgggc 6780
 tcaaccagta cctggctggg tcacagctcc catgtgagcc cgaaccggat gtggcagtgc 6840
 tcacttccat gctcaccgac cctctcata ttacagcaga gacggccaag cgtaggctgg 6900
 ccagggggtc tccccctcc ttggccagct cttcagctag ccagttgtct gcgccttctt 6960
 tgaaggcgac atgtactacc catcatgact ccccgagcg tgacctatc gaggccaacc 7020
 tcctgtggcg gcaggagatg ggcgggaaca taccctgtg ggagtcagaa aataaggtgg 7080
 taatcctgga ctctttcgat ccgattcggg cgggtggagga tgagaggga atatccgtcc 7140
 cggcgagat cctgcgaaaa ccaggaagt tccccccagc gttgcccata tgggcacgcc 7200
 cggattacaa ccctccactg ctagagtcct ggaaggaccc ggactacgtc ccccggtgg 7260
 tacacgggtg ccctttgcca tctaccaagg ccccccaat accacctcca cggaggaaga 7320
 ggacggttgt cctgacagag tccaccgtgt cttctgcctt ggcggagctc gctactaaga 7380
 cctttggcag ctccgggtcg tcggccgttg acagcggcac ggcgactggc cctcccgatc 7440
 aggctccga cgacggcgac aaaggatccg acgttgagtc gtactcctcc atgccccccc 7500
 tcgagggaga gccaggggac cccgacctca gcgacgggtc ttgggtctacc gtgagcgggg 7560
 aagctggtga ggacgtcgtc tgctgctcaa tgcctatac atggacaggt gccttgatca 7620
 cgccatgctc tgcggaggag agcaagttgc ccatcaatcc gttgagcaac tctttgtgc 7680
 gtcaccacag tatggtctac tccacaacat ctgcgagcg aagtctgcg cagaagaagg 7740
 tcaccttga cagactgcaa gtcctggacg accactaccg ggacgtgctc aaggagatga 7800
 aggcgaaggc gtccacagtt aaggctaggc ttctatctat agaggaggcc tgcaaactga 7860
 cgccccaca ttcgccaaa tccaaatttg gctacggggc gaaggacgtc cggagcctat 7920
 ccagcagggc cgtcaaccac atccgctccg tgtgggagga cttgctggaa gacactgaaa 7980
 caccaattga taccaccatc atggcaaaaa atgaggtttt ctgcgtccaa ccagagaaag 8040
 gaggcgcaa gccagctcgc cttatcgtat tccagacct ggggttacgt gtatgcgaga 8100
 agatggccct ttacgacgtg gtctccacc ttctcagc cgtgatgggc ccctcatag 8160
 gattccagta ctctcctggg cagcgggtcg agttcctggt gaatacctgg aatcaaaga 8220
 aatgcctat gggcttctca tatgacaccc gctgctttga ctcaacggtc actgagaatg 8280

acatccgtac tgaggaatca atttaccaat gttgtgactt ggcccccgaa gccaggcagg 8340
 ccataagggtc gctcacagag cggctttatg tcgggggtcc cctgactaat tcgaaggggc 8400
 agaactgcgg ttatcgccgg tgccgcgcaa gtggcgtgct gacgactagc tgccgcaaca 8460
 ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc caggactgca 8520
 cgatgctcgt gaacggagac gaccttgctg ttatctgtga gagtgcggga acccaggagg 8580
 atgcggcggc cctacgagcc ttcacggagg ctatgactag gtattccgcc cccccggg 8640
 acccgcccca accagaatac gacttgagc tgataacgtc atgtccctcc aatgtgtcgg 8700
 tcgcgcacga tgcattccggc aaaaggggtg actacctcac ccgtgacccc accaccccc 8760
 tcgcacgggc tgctggggag acagttagac aactccagt caactcctgg ctaggcaata 8820
 tcatcatgta tgcgccacc ctatgggcga ggatgattct gatgactcat ttcttctcta 8880
 tccttctagc tcaggagcaa cttgaaaaag ccctggattg tcagatctac ggggcctgtt 8940
 actccattga gccacttgac ctacctaga tcattgaacg actccatggt cttagcgcat 9000
 tttcactcca cagttactct ccaggtgaga tcaatagggt ggcttcatgc ctcaggaaac 9060
 ttgggggtacc gcctttgcga gtctggagac atcgggccag aagtgtccgc gctaagctac 9120
 tgtcccaggg ggggagggct gccacttgcg gcaagtacct cttcaactgg gcagtaaaga 9180
 ccaagcttaa actcactcca atcccggctg cgtcccagct agacttgtcc ggctgggtcg 9240
 ttgctgggta caacggggga gacatatatc acagcctgtc tcgtgcccgga ccccgttggt 9300
 tcatgttgtg cctactocta ctttctgtag gggtaggcat ctacctgtc cccaaccggt 9360
 gaacggggag ctaaccactc caggccaata ggccattccc tttttttttt ttc 9413

<210> 18

<211> 328

<212> RNA

<213> Homo sapiens

<400> 18

uugggggcga cacuccacca uagaucacuc ccugugagg aacuacuguc uucacgcaga 60
 aagcgucuaag ccauggcguu aguaugagug uugugcagcc uccaggaccc cccucccgg 120
 gagagccaaua guggucugcg gaaccgguga guacaccgga auugccagga cgaccggguc 180
 cuuuuugga ucaaccgcu caaugccugg agauuugggc gugccccgc gagacugcua 240
 gccgaguagu guugggucgc gaaaggccuu gugguacugc cugauagggg gcuugcgagu 300
 gccccgggag gucucguaga ccgugcau 328

<210> 19

<211> 14

<212> RNA

<213> Homo sapiens

<400> 19

auuugggcgu gccc

14

<210> 20

<211> 27

<212> RNA

<213> Homo sapiens

<400> 20

gccgaguagu guugggucgc gaaaggc

27

<210> 21

<211> 340

<212> DNA

<213> Homo sapiens

<400> 21

atgggcggag ggaagctcat cagtggggcc acgagctgag tgcgtcctgt cactccactc 60

ccatgtccct tgggaaggtc tgagactagg gccagaggcg gccctaacag ggetctccct 120

gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggtcag actgggcagg 180

agatgccgtg gaccccgccc ttcggggagg ggcccggcgg atgcctcctt tgccggagct 240

tggaacagac tcacggccag cgaagtgagt tcaatggctg aggtgaggta ccccgaggg 300

gacctcataa cccaattcag accactctcc tcgcccatt 340

<210> 22

<211> 349

<212> DNA

<213> Homo sapiens

<400> 22

gaggaaagtc cgggctcaca cagtctgaga tgattgtagt gttcgtgctt gatgaaacaa	60
taaatcaagg cattaatttg acggcaatga aatatcctaa gtctttcgat atggatagag	120
taatttgaaa gtgccacagt gacgtagctt ttatagaaat ataaaagggtg gaacgcggta	180
aaccctcga gtgagcaatc caaatttggt aggagcactt gtttaacgga attcaacgta	240
taaacgagac acacttcgcg aaatgaagtg gtgtagacag atggttatca cctgagtacc	300
agtgtgacta gtgcacgtga tgagtacgat ggaacagaac gcggcttat	349

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

<400> 23	
gaagctgacc agacagtgcg cgcttcgtcg tcgtcctctt cgggggagac gggcggaggg	60
gaggaaagtc cgggctccat agggcagggt gccaggtaac gcctgggggg gaaaccacg	120
accagtgcaa cagagagcaa accgcgatg gcccgcgcaa gcgggatcag gtaagggtga	180
aagggtgcgg taagagcgca ccgcgggct ggtaacagtc cgtggcacgg taaactccac	240
ccggagcaag gccaaatagg ggttcataag gtacggcccg tactgaacct gggtaggctg	300
cttgagccag tgagcgattg ctggcctaga tgaatgactg tccacgacag aaccgggctt	360
atcggtcagt ttcacct	377

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

<400> 24	
ccaccgggta cgatcttgcc gaccatggcc ccacaatagg gccggggaga cccggcgta	60
gtggtgggcg gcacggtcag taacgtotgc gcaacacggg gttgactgac gggcaatatc	120
ggctccatag cgtcggccgc ggatacagta aaggagcatt ctgtgacgga aaagacgccc	180
gacgacgtct tcaaacttgc caaggacgag aaggtcgaat atgtcgacgt ccggttctgt	240
gacctgcctg gcatcatgca gcacttcacg attccggctt cggcctttga caagagcgtg	300
tttgacgacg gcttggcctt tgacggctcg tcgattcgcg ggttcagtc gatccacgaa	360
tccgacatgt tgcttcttcc cgatcccag acggcgcgca tcgacctgtt ccgcgcggcc	420

aagacgctga atatcaactt ctttgtgcac gaccggttca ccctggagcc gtactccgc 480
gaccgcgcga acatcgcccg caaggccgag aactacctga tcagcactgg catcgccgac 540
accgcatact tcggcgccga ggccgagttc tacattttcg attcggtgag cttcgactcg 600
cgcgccaacg gtccttcta cgagggtggac gccatctcgg ggtggtggaa caccggcgcg 660
gggaccgagg ccgacggcag tccaaccgg ggctacaagg tccgccacaa gggcgggtat 720
ttccagtggt ccccaacga ccaatacgtc gacctgcgcg acaagatgct gaccaacctg 780
atcaactccg gtttcatcct ggagaaggcg caccacgagg tgggcagcgg cggacaggcc 840
gagatcaact accagttcaa ttcgctgctg cagccgccc acgacatgca gttgtacaag 900
tacatcatca agaaccgc ctggcagaac ggcaaacgg tcacgttcat gcccaagccg 960
ctgttcggcg acaacgggtc cggcatgcac tgtcatcagt cgctgtggaa ggacggggcc 1020
ccgctgatgt acgacgagac gggttatgcc ggtctgtcgg acacggccc tcatcatc 1080
ggcggcctgt tacaccaacg gccgtcgtg ctggccttca ccaaccgac ggtgaactcc 1140
tacaagcggc tggttcccg ttacgaggcc ccgatcaacc tggctatag ccagcgcaac 1200
cggtcggcat gcgtgcgcat ccgatcacc ggagcaacc cgaaggcaa gcggtggag 1260
ttccgaagcc ccgactcgtc gggcaaccg tatctggcgt tctcgccat gctgatggca 1320
ggcctggaag gtatcaagaa caagatcgag ccgagggcg ccgtcgacaa ggatctctac 1380
gagctgcgcg cggaagaggc cgcgagtatc ccgagactc cgaccagct gtcagatgtg 1440
atcgaccgtc tcgaggccga ccacgaatac ctcaccgaag gaggggtgtt cacaacgac 1500
ctgatcgaga cgtggatcag tttcaagcgc gaaaacgaga tcgagccggt caacatccgg 1560
ccgcatccct acgaattcgc gctgtactac gacgtttaag gactcttcgc agtccgggtg 1620
tagagggagc ggcgtgtcgt tgccaggcg ggcgtcgagg tttttcgatg ggtgacggtg 1680
gccggcaacg gcgcgccgac caccgtcgcg aagagcccgt ttaagaacgt tcaaggacgt 1740
ttcagccggg tgccacaacc cgcttgga tcatctccc accgccgagc gggttgtctt 1800
tcacatgcgc cgaaactcaa gccacgtcgt cggccaggcg tgcgtcgcg gccggttcag 1860
gttaagtgtc ggggattcgt cgtgcggcg ggcgtccacg ctgaccaacg gggcagtcga 1920
ctccgaaca ctttgcgcac taccgcctt gcccgccgcg tcaccgtag gtagttgtcc 1980
aggaattccc caccgtcgtc gtttcgccag ccggccgcga ccgcgaccgc attgagctgg 2040
cgcccggtc ccggcagctg gtcggtggcg ttgccgcga ccaacaccag cgcttgcg 2100
gcccgggtg cggtcagcca ggcctgacgg agcagctcca cgtcggtgc ggggaaccaga 2160
tcggcgcccg cgatgacatc cagggattgc agcgtcgagg tgtgtgcag ggcgggaacc 2220
tggtgcgcat gctgtagctg cagcaactgc acggtccatt cgatgtcggc cagtcgcccg 2280
cggcccagtt tgggtgtgtg gttggggtcg gcacgcgcg gcaaccgctc ggactcgata 2340

cgggccttga tgcggcgaat ctgcgcacc gagtacgcgg acacaccgtc gggcggatac 2400
 cgcgttttgt cgaccatccg taggaatcgc tgacccaact cggcatcgcc ggcaaccgcg 2460
 tgtgcgcgta gcagggcctg gatctcccat ggctgtgcc actgctcgta gtatgcggcg 2520
 taggaccca ggggtcggac cagcggaccg ttgcggccct cgggtcgcaa attggcgtcg 2580
 agctccagcg gcggatcgac gctgggtgtc ccagcagcg ccgaaccgc ctcggcgatc 2640
 gatgtcgacc atttcaccgc ccgtgcatcg tcgacgcgg tggccggctc acagacgaac 2700
 atcacgtcgg catccgaccc gtagcccaac tcggcaccac ccagccgacc catgcccgatg 2760
 accgcgatgg ccgccggggc gcgatcgctc tcgggaaggc tggcccggat catgacgtcc 2820
 agcgcggcct gcagcaccgc caccacacac gacgtcaacg cccggcacac ctcggtgacc 2880
 togagcaggc cgagcaggtc cgccgaaccg atgcgggcca gctctcgacg acgcagcgtg 2940
 cgcgcgccgg cgatggcccg ctccgggtcg gggtagcggc tcgccgaggc gatcagcgcc 3000
 cgagccacgg cggcgggctc ggtctcgacg agcttcgggc ccgcaggccc gtctctgtac 3060
 tgttgatga cccgcggcgc gcgcatcaac agatccggca catacgccga ggtaccaag 3120
 acatgcatga gccgcttggc caccgcgggc ttgtcccga gcgtggccag gtaccagctt 3180
 tcggtggcca gcgcctcact gagccgcgg taggccagca gtccgccgtc gggatcgggg 3240
 gcatacgaca tccagtccag cagcctgggc agcagcaccg actgcaccgc tcgcgcggc 3300
 ccgctttgat tgaccaacgc cgacatgtgt ttcaacgcgg tctgcgggtc ctcgtagccc 3360
 agcgcggcca gccggcggc cgcgccctcc aacgtcatgc cgtgggcat ctccaaccgc 3420
 gtcgggccga tcgattccag cagcgggtga tagaagagtt tgggtgtgaa cttogacacc 3480
 cgcacgttct gcttcttgag ttctccgcg agcaccgcgg ccgcacgtt tcggccatcg 3540
 ggccggatgt gggccgcgcg cgccagccag cgcactgcct cctcgtcttc gggatcggga 3600
 agcaggtggg tgcgcttgag ccgctgcaac tgcagtcggt gctcgagcag cctgaggaac 3660
 tcatacgacg cggatcatgt cgccgcgtcc tcacgccga ttagccgcgc ttccgccaac 3720
 gccgccaatg cgtccaccgt ggacgccacc cgtaacgact cgtcgctacg ggcatagaac 3780
 agctgcagta gctgtacggc gaactccacg tcgcgcaatc cgccgctgcc gagtttgagc 3840
 tcgcggccgc ggacatcggc gggcaccagc tgctccacc gccgcgcgat ggcctgcacc 3900
 tcgaccacaa agtcttcgcg ctgcaggtc cgccacacca tcggcatcaa ggcggtcagg 3960
 taacgctcgc caagttccgc gtcgccaacg actggccgtg ctttcagcaa gcgctgaaac 4020
 tcccaggtct tggccagcg ctggtagtag gogatgtgcg actogagcgt acggaccagc 4080
 tcccgttgc gccctccgg acgcagggcg gcgtccacct cgaaaaggc cgccgaggcc 4140
 accgcacatc tctcgtggc cagcgcgcg ttgcgcgggt cggagcgctc ggcaacgaat 4200
 atgacatcga cgtcgtgac gtagttcagt tcgcgcgcac cgcaattgcc catcgcgatg 4260

accgccaggc gcggtggcgg gtgctcgccg cacacgctcg cctcggccac gcgcagcgcc 4320 -
 gccgccagag cggcgtccgc ggcgtccgcc aggcgtgcgg ccaccacggt gaatggcagc 4380
 accggttcgt cctcgaccgt cgcggccagg tcgagagcgg ccagcattag cacgtagtgc 4440
 cgggtactggg ttcgcaatcg gtgcacgagc gagcccggca taccctccga ttcctcgacg 4500
 cactcgacga acgaccgctg cagctggtca tgggacggca gtgtgacctt gccccgcagc 4560
 aatttcagg actgcggatg ggcgaccagg tgatcgccca acgccagcga cgagcccagc 4620
 accgagaaca gccgcccgcg cagactgcgt tcgcgcagca gagccgcgtt gagctcgtcc 4680
 catccggtgt ctggattctc cgacagccgg atcaaggcgc gcagcgcggc atcggcgtcc 4740
 ggagcgcgtg acagcgacca cagcaggtcg acgtgcgcct gatcctcgtg ccgatccac 4800
 ccagctgag ccagacgctc accagcaggg gggtaacta atccgagccg gccaacgctg 4860
 ggcaacttcg gccgctgcgt ggcgagtttg gtcacgacca cgacggtagc gcaaagcgcg 4920
 tcggcgtcgg atcaacoggt agatctgggc tacagcgaca ggtaggtgcg cagctcgtat 4980
 ggcgtgacgt ggctgcggta gttcgccac tccgtgcgt tgttgcgcaa gaaaaagtca 5040
 aaaacgtgct cccccaaggc ctccgcgacg agttcgagg cctccatggc gcgcagcgca 5100
 ctatccaaac tggacggcaa ttctcggtac cccatcgctc ggcgttcctc ggggtgtgagg 5160
 tcccatacgt tgtcctcggc ctgcggggcc agcacgtaac ccttctctac accccgcaat 5220
 cccgcggcca gcagcacggc gaatgtcaga tagggattgc acgccgaatc agggctgcgt 5280
 acttcgaccc gccgcgacga ggtcttgtgc ggcgtgtaca tcggcaccgc cactagggcg 5340
 gatcggttgg cgcccccca cgacgcggcc gtgggcgctt cgcgcacctg caccagccgc 5400
 ttgtaagagt tgaccactg atttgtgacc gcgctgatct cgcaagcgtg ctccaggatc 5460
 ccggcgatga acgatttacc cacttccgac agctgcagcg gatcatcagc gctgtggaac 5520
 gcgttgacat caccctcgaa caggctcatg tgggtgtgca tcgccgagcc cgggtgtggtg 5580
 ccgaatggct tgggcatgaa cgacgcccgg gcgcctctt ccagcgcgac ttctttgatg 5640
 acgtacggga aggtcatcac gttgtcagcc atcgacagag cgtcggcaaa ccgcaggtcg 5700
 atctcctgct ggccgggtgc gccttcgtga tggctgaact ccaccgagat gcccatgaat 5760
 tccagggcat cgatcgctg gcggcgaaag ttcaaggcgg agtcgtgcac cgcttggtcg 5820
 aaatagccgg cgttgtcgac cgggacgggc accgaccgt cctcgggtcc gggcttgagc 5880
 aggaagaact cgatttcggg atgcacgtag caggagaagc cgagttcgcc ggccttcgtc 5940
 agctgcogcc gcaacacgtg ccgcgggtcc gccacgacg gcgagccgtc cggcatggtg 6000
 atgtcgcaaa acatccgcgc tgagtgtggtg tggccggaac tggtggccca gggcagcacc 6060
 tggaaggctg acgggtccgg gtgcgcccac gtatcggatt ccgagaccgc cgcaaagccc 6120
 tcgatcgagg atccgtcgaa gccgatgcct tcctcgaagg cgccctcgag ttcggctggg 6180

gcgatggcga ccgacttgag gaaaccgagc acgtctgtga accacagccg gacgaagcgg 6240
atgtcgcgtt cttccagggt acgaagaacg aattccttct gtcgggtccat acctcgaaca 6300
gtatgcaotg tctgttaaaa cegtgttacc gatgcccggc cagaagcgtt gcggggcggc 6360
ccgcaagggg agtgcgcggt gagttcaggg cgcgcaccgc agactcgtcg gcggcaaggt 6420
cccgtcgaga aaatagtgca tcaccgcaga gtccacacac tggttgccat cgaacaccgc 6480
agtgtgttg gtgccgtcga aggtgatcag cgggtgcgcc agctggcggg ccagggtctac 6540
cccggactga tacggagtgg ccgggtcgtg ggtggtggac accacgacga ccttgccagc 6600
cccggccggc gccgcggggg gggcgctcga cgttgccggc accggccaca gcgcgcacag 6660
atcgcggggg gcggatccgg tgaactgcc gtagctaagg aacggggcga cctgacggat 6720
ccgttggtcg ggggccaccc aggcgcgtg atcggccggg gtgggcgcat cgacgcaccg 6780
gaccgcgttg aacgcgtcct ggtcgttgct gtagtgcccg tctgcatccc ggccgtcata 6840
gtcgtcggca agcaccagca agtcgccggc gtcgctgccg cgtgcagcc ccagcagacc 6900
actggtcagg tacttccagc gctgagggt gtacagcgcg ttgatggtgc ccgtcgtcgc 6960
gtcggcgtag ctcaggccac gtggatccga cgtcttacc ggcttctgca ccagcgggtc 7020
aaccagggcg tggtagcggg tgaccactg ggccgagtcg gtgccagag ggcaggccgg 7080
cgagcggcg cagtcggcg cgtagtcatt gaaagcggc tgaaatccc ccatttggt 7140
gatgctttcc tcgattggg taacggctg atcgatagc ccgtcgagga ccacgcccg 7200
cacatgagta ccgaaccgtt ccaggtaagc ggtgccaac tcggtgcgt agctgtatcc 7260
gaggtagttg atctgatcgt cacctaacgc ttggcgaacc atgtccatgt cccgtgcgac 7320
ggacgcggta ccgatattg ccaagaagct gaagccatc cggccaacac agtcctgggc 7380
caactgccgg tagacctgt cgacgtggg gacaccggc ggactgtagt cggccatcgg 7440
atcgcgccgg tacgcgtcga actcggcgtc ggtgcgacac cgcaacgcag gggtcgagtg 7500
gccgaccct ctcgggtcga agcccaccg gtcgaagtgg cggaagtgt cgggtgcggc 7560
gatcgcgggt gccatagcgg cgaccatgtc gaccgcgcg gcccgggtc cccaggatt 7620
gaccagcagt gctccgaatc gctgtcccgt cgcggggacg cggatcacg ccaacttcgc 7680
ttgtgtcca ccgggttggt cgtagtcgac ggggacggac accgtcgcgc agcgtgcagt 7740
gcgaatttcg ctggtgtcgg cgatgaactc gcggcagctg ttccaactct gttgcggcgc 7800
cacgaccggc gcaccggggg ttggccggc gccgggttct tcagtcgcgc cggccaacgg 7860
gggcgctgct aggggcagtc cgcgagcag caaccogaag gacagcagc ccgagctcaa 7920
cggctcgcgg cgccacatgg ccgccatcgt ctcaccggcg aatacctgtg acggcgcgaa 7980
atgatcacac cttcgtttct tcgccccgt agcacttggc gccgtgggc ggcgtggtgc 8040
cgccgattaa atacgccgtc acgtactcgt caatgcagct gtcgccctg aataccaccg 8100

tgtgctgggt tccgtcgaag gtcagcaacg aaccgcgaag ctggttcgcc aggtcgaccc 8160
 cggccttgta cggcgtcgcc gggatcatggg tggtaggatac caccaccgtc ggcactaggc 8220
 cgggcgcgca gacggcatgg ggtgacttg tgggtggcac cggccagaac ggcaggtgc 8280
 ccagcggcgc atcaccggtg aacttcccgt agctcatgaa cggtagcatc tcccgggcgc 8340
 ggcggtcttc gtcgatgacc ttgtcgcgat cggtaacccg gggctgatcg acgcaattga 8400
 tcgccacccg cgcgtcaccg gaattgttgt agcggccgtg cgagtcccga cgcattgaca 8460
 tgtcggccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgtcgg 8520
 tcaagtgttg ccacagattc ggtgagtaca gcgccataat ggtgcccacg atggcgtcgc 8580
 tataactcag cccgcgcgga tccttcgtgc gcgccggcct gctgatcctc gggttgtccg 8640
 ggtcgaccaa cggatcgacc aggtgtggt agacctcgac ggctttggcc gggtcggcgc 8700
 ccagcgggca gccgcggttc ttggcgcagt cggcggcata gttgttgaa gcgtcctgga 8760
 agcccttggc ctggcgcagc tccgcctcga tgggatcggc attggggtcg acggcaccgt 8820
 cgagaatcat tgccgcacc cgtcgcgaa attcctcggc atacgcggag ccgatccggg 8880
 tgccgtacga gtagccaggt taggtcagct tgtcgtcgcc caacgccgcg cgaatggcat 8940
 ccaggtcctt ggcgacgttg accgtcccga catgggccag aaagttcttg cccatcttgt 9000
 ccacacagcg accgacgaat tgcttgggtc cgttctcgat gtgcgccaca ccctcccggc 9060
 tgtagtcaac ctgcggctcg gccgcagcc ggtcgttgtc ggcacggag ttgcaccaga 9120
 tcgccggccg ggacgacgc accccgcggg ggtcgaaccc aaccaggtcg aacctttcgt 9180
 gcacccgctt cggcaatgtc tggaagacgc ccaaggcggc ctcgataccg gattcgccgg 9240
 gtccaccggg atttatgacc agcgaaccga tcttgtctcc cgtcgcggga aagcgaatca 9300
 gcgcagcgc cgcacgtca ccatcggggc ggtcgtagtc gaccggtaca gcgagcttgc 9360
 cgcataacgc gccgcgggg atctttactt gcgggtttga cgaccggcac ggtgtccact 9420
 ccacgggtg gccagcttc ggctccgcca tacgagcgcg tccccgacc acgcggatgc 9480
 agcccacaag aaccaacgcc acggcggcga gcgcggccca gatcaacagc atgcgcgcga 9540
 tcttgtcgcg gcgagacagc ctcatgccca caatgctgcc agagcagacc cgagatcctg 9600
 gccagcggcc accgtcggcc gactaaccgg ccgctgccag cagtctgcc atcgccgatg 9660
 gcgaactcgt cggccatccc ccatacgtcc ggtaacagat ccgggcaaga caccgacccg 9720
 tcgaccggat ccggcacggg cgcgtcggcc tcggcgggtg acaactgcga catcagggtg 9780
 gcgctggcac cccgtccacg ccggcatggt gcaccttggc catcgcccga gggcgatccc 9840
 cgatgccgtc cacccttcg acgaacccat ctcccacggc ggtcgcgggc agcgacgcga 9900
 tgtggcgcga gatctccgag agttcggccc gccgcggcg cgacggcaac ccgatgccgt 9960
 gcaagtgcag atcgatgtga ggttcaaggt tcagcgcact gctggcaagc tttttccgaa 10020

accgcggcct cgccttgatc tggagtcaga acgcgtcacg cagccggtca aaggcgtaac 10080
ccatgctcga gcaaaccatgc atgggctgag tggacgtttc cagacacagc aactggcgtc 10140
caggccactg agccgctgca tgcgcgatgg tatgccgatg ggggccccgg gcgcgtctga 10200
ggggaagaag tggcagactg tcagggtccg acgaaccggg ggaccctaac gggccacgag 10260
gatcgaccgg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320
ccccggagg ctccggggcg cggaccaaga tccgcacca ccacctacag agatggaagg 10380
ccgacggcca caagtggggc atgtgacgg cctacgacta ttcgacggcc cggtatcttcg 10440
acgaggccgg catcccgtg ctgctggctg gtgattcggc ggccaacgtc gtgtacggct 10500
acgacaccac cgtgccgatc tccatcgacg agctgatccc gctggtcggt ggcgtggtgc 10560
ggggtgcccc gcacgcactg gtcgtcgccg acctgcogtt cggcagctac gaggcggggc 10620
ccaccgccgc gttggccgcc gccaccggg tcctcaagga cggcggcgca catgcggtca 10680
agctcgaggg cgggtgagcg gtggccgagc aaatcgctg tctgaccgg gcgggcatcc 10740
cgggtgatggc acacatcggc ttcacccgc aaagcgtaa caccttgggc ggcttcgggg 10800
tgcagggccg cggcgacgcc gccgaacaaa ccatcgccga cgcgatgcc gtcgccgaag 10860
ccggagcggt tgcgcgtctg atggagatgg tgcccgcga gttggccacc cagatcaccg 10920
gcaagcttac cattccgacg gtcgggatcg gcgctgggcc caactgcgac ggccagggtcc 10980
tggtatggca ggacatggcc gggttcagcg gcgccaagac cggccgcttc gtcaaacggt 11040
atgccgatgt cgggtggtgaa ctacgcogtg ctgcaatgca atacgccaa gaggtggccg 11100
gcggggtatt ccccgctgac gaacacagtt tctgaccaag ccgaatcagc ccgatgcgcg 11160
ggcattgcgg tggcgccctg gatgccgtcg acgccggatt gccggcgcg acgcgccagc 11220
gggacccatc ggcgtcgctg tcgcccgttg agcccgggt gagcccagac attcgatgtg 11280
cccaacacca tcgccacag cccaattgat gtggcactct atgcatgcct atccccgacc 11340
aaccaccacc gggcgacgc atcatgaccg gaggcgaaga tgccagtaga ggcgccaga 11400
ccagcgcgcc atctggaggt cgagcgcaag ttcgacgtga tcgagtcgac ggtgtcgccg 11460
tcgttcgagg gcatcgccgc ggtggttcgc gtcgagcagt cggcgacca gcagctcgac 11520
gcggtgtact tcgacacacc gtcgcacgac ctggcgcgca accagatcac cttgcggcg 11580
cgcaccggcg gcgcgcgac cggctggcat ctgaagctgc cggccggacc cgacaagcg 11640
accgagatgc gagcaccgct gtccgcatca ggcgacgctg tgccggccga gttgttgat 11700
gtggtgctgg cgatcgccg cgaccagccg gttcagccgg tcgcgcggt cagcactcac 11760
cgcgaaagcc agatcctgta cggcgccggg ggcgacgcgc tggcggaatt ctgcaacgac 11820
gacgtcaccg catggtcggc cggggcatc cagccgctg gtgcagcgga caacggccct 11880
gccgaacagc agtggcgoga atgggaactg gaactggtca ccacggatgg gaccgccgat 11940

accaagctac tggaccggct agccaaccgg ctgctcgatg ccggtgccgc acctgccggc 12000
cacggctcca aactggcgcg ggtgctcggg gcgacctctc ccggtgagct gcccaacggc 12060
ccgcagccgc cggcggatcc agtacaccgc gcggtgtccg agcaagtcga gcagctgctg 12120
ctgtgggatc gggccgtgcg ggccgacgcc tatgacgccg tgcaccagat gcgagtgcg 12180
acccgcaaga tccgcagctt gctgacggat tcccaggagt cgtttggcct gaaggaaagt 12240
gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tgggcgtagc ccgggacgcc 12300
gaggtactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcgccgga gctggtacgc 12360
ggccgggtgc gcgagcgctt ggtagacggg gcgcggcgcc gataccagac cgggctgcgg 12420
cgatcactga tcgcattgcg gtgcgacggg tacttccgtc tgctcgacgc tctagacgcg 12480
cttgtgtccg aacgcgcccc tgccacttct ggggaggaat cggcaccggg aaccatcgat 12540
gcggcctacc ggcgagtcg caaagccgca aaagccgcaa agaccgccg cgaccaggcg 12600
ggcgaccacc accgcgacga ggcattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660
tacaccgcg cggtactgg ggcggacaat gtgtcacaag aagccaaggt catccagacg 12720
ttgctaggcg atcatcaaga cagcgtggtc agccgggaac atctgatcca gcaggccata 12780
gccgcgaaca ccgccggcg gaacacctc acctacggtc tgctctacca acaggaagcc 12840
gacttggccg agcgtgccc ggagcagctt gaagccgcgc tgcgcaaact cgacaaggcg 12900
gtccgcaaag cacgggattg agcccgccag gggcgacga gttggcctgt aagccggatt 12960
ctgttccgcg ccgccacagc caagctaacg gcggcacggc ggcgaccatc catctggaca 13020
cacggttacc ggtgacctg agcggcctac ccgcaggctc ggcgagcaa ccctcaagcg 13080
cctgcgcggc cgcactttcg gtgcggcctt cttggccttg cttcgggttg ggtttgccta 13140
gccaccccgg tcacccggaa tgctggtgcg ctcttaccgc accgtttcac ccttgccacc 13200
acgaggatgg cggctctgtt tctgtggcac tttcccgca gtcacctcg attgccgtta 13260
gcaatcccc tgcctgtga agtcggact ttctcgact cgacgctgaa cctcgtgaat 13320
ccacacaagc cctacgcgag ccgcggccgc ccagccaact catccgcgac gaccacgcta 13380
ccccgctggg cgggtgcgcg gccagtgtga ccgctggacg acacggctag tcggacagcc 13440
gatccggcgg gcagtcctta tcgtggactg gtgacacggt gggacaaacg cgtcgactcc 13500
ggcgactggg acgccatcgc tgccgaggtc agcgagtacg gtggcgact gctacctcg 13560
ctgatcccc ccggcgaggc cggccggctg cgcaagctgt acgcgacga cggcctgttt 13620
cgctcgacgg tcgatatggc atccaagcgg tacggcgccg ggcagtatcg atatttccat 13680
gccccctatc ccgagtgate gagcgtctca agcaggcgct gtatcccaa ctgctgccga 13740
tagcgcgcaa ctggtgggcc aaactgggcc gggaggcgcc ctggccagac agccttgatg 13800
actggttggc gagctgtcat gccgccggcc aaaccgatc cacagcgctg atgttgaagt 13860

acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgtttccgc 13920
tgcagggtgt gatcaacctg agcgatccgg aaaccgacta caccggcggc gagttcctgc 13980
ttgtcgaaca gcggcctcgc gcccaatccc ggggtaccgc aatgcaactt ccgcagggac 14040
atggttatgt gttcacgacc cgtgatcggc cggtgccgac tagccgtggc tggtcggcat 14100
ctccagtgcg ccatgggctt tcgactattc gttccggcga acgctatgcc atggggctga 14160
tctttcacga cgcagcctga ttgacgccca tctatagata gcctgtctga ttcaccaatc 14220
gcaccgacga tgcccatcgc gcgtagaact cggcgatgct cagcgatgcc agatcaagat 14280
gcaaccgata taggacgccc gaccggcat ccaacgccag ccgcaacaac attttgatcg 14340
gcgtgacatg tgacaccacc agcacgctgc cgccttcgta gccaacgatg atccgatcac 14400
gtccccgccg aaccgcgcgc agcacgctgc cgaagctttc cccaccggg ggcgtgatgc 14460
tggtgtcctg cagccagcga cggtgacgct cgggatcgcg ttctgcggcc tccgcgaacg 14520
tcagccccctc ccaggcgccg aagtgggtct cgaccaggtc gtcacgacg accacgtcca 14580
gggccagggc tctggcgggc gtcaccgcgc tgcgtaagc ccgctgtagc ggcgaggaga 14640
ccaccgcagc gatcccgccg cgcgcgcga gataccggc cgcgcacca acctggcgcc 14700
acccacctc gttcaacccc gggttgcgc gccccaata gcggcgttgc tccgacagct 14760
ccgtctgccc gtggcgcaac aaaagtagtc ggggtgggtg accgcgggcg ccggtccagc 14820
cgggagatgt cggtgactgc gtcgcaacga ttttggcagg atccgcatcc gccgcagccg 14880
attgcgcggc ggcgctccatc gcgtcattgg ccaaccggtc tgcatacgtg ttccgggcac 14940
gcggaacca ctcgtagttg atcctgcgaa actgggacgc caacgcctga gcctggacat 15000
agagcttcag cagatccggg tgettgaact tccaccgcc ggacatctgc tccaccacca 15060
gcttgagtc catcagcacc gcggcctcgc tggcacctag tttcacggcg tcgtccaaac 15120
cggctatcag gccgcggtat tcggcgacgt tgttcgtcgc ccggccgac gcctgcttgg 15180
actcgccag cagggtggag tgatcggcg tccacaccac cgcgcggtat ccggccggtc 15240
cgggattgcc ccgcatccg ccgtcggctt cgatgacaac tttactcct caaatcctc 15300
gagccgcaac aagatcgctc cgcattccgg gcagcgcaac acttcacct ccggcgccgc 15360
cgagatctgg gccagctcgc cgcggccgat ctcgatccgg caggcaccac atcgatgacc 15420
ttgcaaccgc ccggcccctg gcccgctcc gggcgctgt ctttcgtaga gcccgcgaag 15480
ctcgggatca agtgctgcgc tcagcatgtc gcgttcgat gaatgttgg gcccggctt 15540
gtcgatttcg gcaagtgcct cgtccaaagc ctgctgggcg gcggccaggc oggcccgcaa 15600
cgcttgagc gcccgcgact cggcggtctg ttgagcctgc agctcctcgc ggcttccag 15660
cacctccagc agggcatctt ccaaactggc ttgacggcgt tgcaagctgt cgagctcgtg 15720
ctgcagatca gccaatgtct tggcgtccgt tgcaccgaa gtgagcaacg accggtccc 15780

gtcgccacgc ttacgcaccg catcgatctc cgactcaaaa cgcgacacct ggccgtccaa 15840
gtcctccgcc gcgattcgca gggccgccat cctgtcgttg gggcggttg gctcggcctg 15900
cacctgctgg taagccgccc gctgcggcag atgggtagcc cgatgcgcga tccgggtcag 15960
ctcagcatcc agcttcgcca attccagtag cgaccgttgc tgtgccactc cggctttcat 16020
gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cgggtgcagat ggtgcacaca 16080
cgcaccggca gcgacgcgcc gaaatgagac cgcaacactt cggcggcctg gccgcaccac 16140
gggaattcgc ttgccaatg cgcgacgtcg atcagggccca cttgcgaagc tcggcaatgc 16200
tcgtcggctg gatgatgtcg cagatcggcc gtaacgtacg cttgcacgtc cgcggcggcc 16260
acggtggcaa gcaacgagtc cccggcgccg ccgcagaccg cgaccgcga caccagcagg 16320
tcgggatccc cggcggcgcg cacaccggtc gcagtcggcg gcaacgcggc ctccagacgg 16380
gcaacaaagg tgcgcagcgg ttccgggtttt ggcagtctgc caatccggcc taaccgcgtg 16440
ccgaccggcg gtggtaccag cgcgaagatg tcgaatgccg gctcctcgta agggtgcgcg 16500
gcgcgcacgc ccgccaacac ctccggcgcg gctcgtcggg gtgcgacgac ctcgaccggg 16560
tcctcggcca cccgttcgac ggtaccgacg ctgcctatgg cgggcgacgc cccgtcgtgc 16620
gccaggaact gcccggtacc cgcgacactc cagctgcagt gcgagtagtc gccgatatgg 16680
ccggcaccgg cctcaaagac cgctgcccg accgcctctg agttctcgcg cggcacatag 16740
atgaccact tgtcgagatc ggccgctccg ggcaccgggt cgagaacggc gtcgacggtc 16800
agaccaacag cgtgtgccag cgcgtcggac acaccggcg acgcgcagtc ggcgttggtg 16860
tgcgcggtaa acaacgagcg accggtccgg atcaggcggg gcaccagcac accctttggc 16920
gtggttgccg cgaccgtatc gacccacgc agtaacaacg ggtggtgcac caatagcagt 16980
ccggcctggg gaacctggtc caccaccgcc ggcgtcgcgt ccaccgcaac ggtcacgaa 17040
tccaccacgt cgtcggggtc gccgcacacc agaaccaccg aatccacga ctgggcaagc 17100
cgcggcgggt aggcctggtc cagcacgtcg atgacatcg ccagccgcac actcatcggc 17160
gtcctccacg ctttgccac tcggcgatcg ccgccaccag caccggccac tccgggcgca 17220
ccgcgcgccg caggtaccgc gcgtccaggc cgacgaaggt gtcaccgcgg cgcaccgcaa 17280
ttcctttgct ctgcaaatag ttctgtaatc cgtcagcatc ggcgatgttg aacagtacga 17340
aaggggccc accatcgacc acctcggcac ccaccgatct cagtcgggcc accatctccg 17400
cgcgcagcgc cgtcaaccgc accgcatcgg ctgcggcagc ggcgaccgcc cggggggcgc 17460
agcaagcagc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgct cgctgcacgg 17520
tcaaccgagc cagcacgtct ggcgagccga gcgcgtagcc caccgcgaat ccggccagcg 17580
accacgtttt cgtcaagcta cggagacca gcacatcggg cagcgagtca tcggccaacg 17640
attgcggctc gccgggaacc caatcagcga acgcctcgtc gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcagc tgctcgcgga ggtgcagcac cgagggtggg ttggtcggat 17760
taccacgac gacaaggctg gcgtcgtcag gcacgtgcgc ggtgtccagc acgaacggcg 17820
gctttaggac aacatggtgc gccgtgattc cggcagcgct caaggctatg gccggctcgg 17880
tgaacgcggg cacgacgatt gctgcccgca ccggacttag gttgtgcagc aatgcgaatc 17940
cctccgccgc cccgacgagc gggagcaact cgtcacgggt tctgccatga cgttcagcga 18000
ccgctctttg cggccgggtg acatcgtcgg tgctcggata gggggccagc tccggcagca 18060
gcgcggcgag ctgccggacc aaccattccg ggggcccgtc atggcggacg ttgacggcga 18120
agtccagcac gccgggcgog acatcctgat caccgtggta gcgcgccgcg gcaagcgggc 18180
tagtgtctag actcgccaca gcgtcaaaca gtagtgggcc ggtgtgcggg ccaagaatcc 18240
agagcaccgc cgacgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag 18300
ggcgtcggcg gtgatgatcg tcaggccaag cagctgtgcc tgggcgatga gcacacggtc 18360
gaatggatgt cgatgggat cgggaagctc tcgggtgcgc agtgtgtgcg tggtaactg 18420
acagcggcga cgtgccgcag cggcgcatte gatcgggcac gtaagaagcc gatggctcgg 18480
gcggcgggag cttgccgagg cggtagttga tcgcgatctc ccaggcactg gcggccgaca 18540
agagaatgct gttgcggaog tctgaacaa tcgcccgtgt ttcgttgacg gcattccgag 18600
ccaaacgtgg gtgtcgatga ggtagcgctt caccggtgaa agcgttcgag cacgtcgtct 18660
gacaacggag cgtccaaatc gtcgggcacg cggtaacgc catggtcaat gcctaaccgc 18720
cgagtctcat gaggatgcag cggcacaagc tttgctaccg gctcgccgcg gcgggcaatc 18780
tcaacctctg cccgccgtag acgagccgca gcagctcgga caggcgtgtc ttcgcctcgt 18840
gaacccgac ccgcttcgca gggcccaga ctttcgcgtc gaccacctgc tcaccaaact 18900
tcgcgatcat cgcctgatac cacagcgcca acgggtagcg gtttgtcaa ccgcttcgtc 18960
aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaattgccc gcctcctoca 19020
cgcccgccg caccggcgcg atcgtcgccg ggtgaatcgc cgcagctggt gatcttcgat 19080
ctggacggca cgctgaccga ctggcgcgcg ggaatcgtat ccagcttccg acacgcgctc 19140
aaccacatcg gtgcccagc acccgaaggc gacctggcca ctcacatcgt cggcccgcgc 19200
atgcatgaga cgctgcgcgc catggggctc ggcgaatccg ccgaggaggc gatcgtagcc 19260
taccgggccc actacagcgc ccgcggttgg gcgatgaaca gcttgctcga cgggatcggg 19320
ccgctgctgg ccgacctgc caccgcgggt gtcggctgg ccgtcgccac ctccaaggca 19380
gagccgaccg caccggcaat cctgcgccac ttcggaattg agcagcactt cgaggctatc 19440
gcgggcgcga gcaccgatgg ctgcgaggc agcaaggctc acgtgctggc ccacgcgctc 19500
gcgcagctgc ggccgctacc cgagcgggtg gtgatggtcg gcgaccgcag ccacgacgctc 19560
gacggggcgg ccgcgcacgg catcgacacg gtggtggtcg gctggggcta cgggcgcgcc 19620

gactttatcg acaagacctc caccaccgtc gtgacgcacg cgcaccgat tgacgagctg 19680
agggaggcgc taggtgtctg atccgctgca cgtcacattc gtttgtacgg gcaacatctg 19740
ccggtcgcca atggccgaga agatgttcgc ccaacagctt cgcaccgtg gcctgggtga 19800
cgcggtgcca gtgaccagtg cgggcaccgg gaactggcat gtaggcagtt gcgcccacga 19860
gcgggcgggc ggggtgttgc gagcccacgg ctaccctacc gaccaccggg ccgcacaagt 19920
cggcaccgaa cacctggcgg cagacctgtt ggtggccttg gaccgcaacc acgctcggct 19980
gttgcggcag ctccgctcgc aagccgcccg ggtacggatg ctgcggtcat tcgaccacg 20040
ctcgggaacc catgcgctcg atgtcgagga tccctactat ggcgatcact ccgacttcga 20100
ggaggtcttc gcggtcatcg aatccgcctt gcccgccctg cagcactggg tcgacgaacg 20160
tctcgcgcgg aacggaccga gttgatgcc cgcctagcgt tcctgctgcg gcccggtcgg 20220
ctggcggttg ccctggctcg ggtcgcgttc acctacctgt gctttacggg gctcgcgcgg 20280
tggcagctgg gcaagaatgc caaaacgtca cgagagaacc agcagatcag gtattccctc 20340
gacacccgcg cggttccgct gaaaaccctt ctaccacagc aggattcgtc ggcgccggac 20400
gcgcagtggc gccgggtgac ggcaaccgga cagtacctc cggacgtgca ggtgctggcc 20460
cgactgcgcg tgggtggagg ggaccaggcg tttgaggtgt tggccccatt cgtggtcgac 20520
ggcggaacca ccgtcctggt cgaccgtgga tacgtgcggc cccaggtggg ctgcacgta 20580
ccaccgatcc ccgcctgcc ggtgcagacg gtgaccatca ccgcgcggct gcgtgactcc 20640
gaaccgagcg tggcgggcaa agaccattc gtcagagacg gcttcagca ggtgtattcg 20700
atcaataccg gacaggtcgc cgcgctgacc ggagtccagc tggctgggtc ctatctgcag 20760
ttgatcgaag accaaccgga cgggctcggc gtgctcggcg ttccgcatct agatcccggg 20820
ccgttcctgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggcttg 20880
ggctatttcg cctacgccga gatccggggc cgcgcgggg aaaaagcggg gtcgccacca 20940
ccggacaagc caatgacggt cgagcagaaa ctgcgtgacc gctacggccg ccggcggtaa 21000
accaacatca cggccaatac cgagccccc gcctggacca cccgcgacag caccacggcg 21060
cggcgagat cggccacctt gggcgaccgg ccgtcgcca aggtggggcg gatctgaac 21120
tcatggtggt accgggtggg cccaccacgc cgcacgtcaa gcgcccagc aaacgccgcc 21180
tcgacgacac cggcggtggg gctgggatgg cgggcccgt cgcgccgcca ggcccgtacc 21240
gcaccgcggg gcgaccacc gaccaccggc gcgcagatca ccaccagcac cgcgctcgcc 21300
cgtgcgcaa catagtggc ccagtcattc aatcgtgctg cagcccaacc gaatcgagga 21360
taacgcggcg agcggtagcc gatcatcgag tccagggtgt tgatggcacg atatcccagc 21420
accgcaggca cgcgctcga agccgccac agcagcggca ccacctggg gtcggcggtg 21480
tttcggcca ccgactccag cgcggcacgc gtcaggcccg ggccgcccag ctgggcccgg 21540

tcacgcccgc acagegacgg cagcagccgt cgcgccgcct cgacatcgtc gcgctccaac 21600
agggtccgata tctggcggcc ggtgcgcgcc agcgaagttc cgcccagcgc tgcccagggtg 21660
gccgtcgcgg tggccgccac gggccaggac ctgcgggta gccgctgcag tgccgcgccg 21720
agcaagccca ccgcgccgac cagcaggccg acgtgtaccg caccggcgac ccggccgtca 21780
cggtaggtga tctgctccag cttggcggcc gcccgaccga acagggccac cggatgacct 21840
cgtttggggg cgccgaacac gacgtcgagc aggcagccga tcagcacgcc gacggccctg 21900
gtctgccagg tcgatgcaa cactccggca gcgtcgaca cgtggtctac gctcagctat 21960
ttatgacctc atacggcagc tatccacgat gaagcggcca gctaccggg ttgccgacct 22020
gttgaaccgg gcggcaatgt tggtgccggc agcgaatgtc atcatgcagc tggcagtgcc 22080
gggtgtcggg tatggcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22140
gttcaagcgg gcccgacca ccggcaccta cctggcgggtg gcgaccatcg ggacggaatc 22200
cgaccgagcg ctgatccggg gtgccgtgga cgtcgcgcac cggcaggttc ggtcgacggc 22260
ctcgagccca gtgtcctata acgccttcga ccgaagttg cagctgtggg tggcggcgtg 22320
tctgtaccgc tacttcgtgg accagcacga gtttctgtac ggcccactcg aagatgccac 22380
cgccgacgcc gtctaccaag acgcaaacy gttagggacc acgtgcagg tgccggaggg 22440
gatgtggccg ccggaccggg tcgcgttcga cgagtactgg aagcgctcgc ttgatgggct 22500
gcagatcgac gcgccggtgc gcgagcatct tcgcgggggtg gcctcggtag cgtttctccc 22560
gtggccgttg cgcgcggtgg ccgggcccgt caacctgttt gcgacgacgg gattcttggc 22620
accggagttc cgcgcgatga tgcagctgga gtggtcacag gccagcagc gtcgcttcga 22680
gtggttactt tccgtgctac ggttagccga ccggctgatt ccgcacggg cctggatctt 22740
cgtttaccag ctttacttgt gggacatgcg gtttcgcgcc cgacaaggcc gccgaatcgt 22800
ctgatagagc ccggccgagt gtgagcctga cagcccgaca ccggcggcgt gtgtcgcgtc 22860
gccaggttca cgctcggcga tctagagccg ccgaaaacct acttctgggt tgcctccga 22920
atcaacgtgc tgatctgctc gagcagctca cgcatatcgg cgcgcacgc atccaccgcg 22980
gcatacaggt cggccttggg ccgggcgagc tggtcgacg tcattggccg caccggcggg 23040
gctgtctgtc gcgcgcgct gtcgctttga aaccagggtc gtcacccac gaccacgaca 23100
ctgccatata cggcgccccg ccgacaacga agcacagcta gccggtgggc gcggacggga 23160
tcgaaccgcc gaccgctggt gtgtaaaacc agagctctac cgctgagcta cgcgcccag 23220
accgccgag gctacacgcc ttgcggccaa gcacccaaaa ccttagggcg taagcgcgcg 23280
cagagcgtcg gtccacagcc gctgatcgcg aacttcaccc ggctgcttca tctcggcgaa 23340
ccgaatgatc cctgaccgat cgaccacaaa ggtgccccgg ttagcgatgc cggcctgctc 23400
gttgaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtcg acaacagcgg 23460

aaacgtgaat ccgctctgcg tcgcccagat cttgtgagtg ggtggcgggc ccaccgaaat 23520
cgctagcgcg gcgctgtcgt cgttctcaaa ctcgggcagg tgatcacgca actggtccag 23580
ctcgccctgg cagatgcccg tgaacgcaa cggaaagaac accaacagca cgttctttgc 23640
accccggtag ccgcgagggg tgacaagctg ctgattctgg tcgcgcaacg tgaagtcagg 23700
ggcgggtggc cggacgttca gcatcagcgc ttgccagccc gcgatttcgg ctgtaccaat 23760
ctgctggcgc tccagttgcc cagattgacc gacgaggtcg gcatcagccc agctgtgggc 23820
gccgcctcgg caatctcggc gggcaataca tggccgggct ggccggtctt gggcgtcacc 23880
acccaaatca caccgtcctc ggcgagcggg ccgatcgcat ccatcagggt gtccacaaaa 23940
tcgccgtcgc catcacgcca ccacaacagg acgacatcga tgacctcgtc ggtgtcttca 24000
tcgagcaact ctccccgca cgcttcttcg atggccgcgc ggatgtcgtc gtcggtgtct 24060
tcgtcccagc cccattcctg gataagttgg tctcgttgga tgcccaattt gcgggcgtag 24120
ttcgaggcgt gatccgccgc gaccaccgtg gaacctcctt cagtctcgc gggccatgtg 24180
cacaccgtcg cgatgggcat tatcgtcgca cagccagaac cggtcacccc gccgcctca 24240
gaaggcggcc acgcacattg tcaatgcctt tgtcttggtg tcgttgagcc gatcaacccg 24300
ccggttgaat tccgctgtcg acgcgtgcgc accgatggca tttgccaccg cgcgggcgc 24360
gtcgacatat gcgttgagcg catccccag ttgcgcggac agcgcggcgc tcagactgcc 24420
tgagaccgtc gaggcactgt tgttgagcgc gtgatggcc ggaccttcgg tcggcccgg 24480
gttgcgcccc tgattgaacg cggccacgta ggcgttcacc ttgtcgatgg cgtccttgt 24540
ggtggccgcc agcgcgtcac acgaggtgag aatcgcttg gtcgtcagcg attgttgcg 24600
ctgcgactcc cggatgctcg acgtcgccgc cgaagccgac accgacgcgg acaccgacga 24660
gcggtaggcc ggtgcgacgt tgggtgtcgg catggccgta ccgtcgggtga cagtggatca 24720
tccgacgac cccatcagca gcagcgcgat gcagccgagc gccagggcgc ctgcctggg 24780
gagctcccc ccgtgcctgc gaggcacggc gcgccatccg atgagcacgg catgtgaggt 24840
tacctggctg cagcgcgacc gcgctggccg tgggtgtgtc cgcatccgca gaaccgagcg 24900
gagtgcggct atccgcgcc gacgccggtg cggcacgata gggggacgac catctaaaca 24960
gcacgcaagc ggaagcccgc cacctacagg agtagtgctg tgaccaccga tttcgccgc 25020
cacgatctgg cccaaaactc aaacagcgca agcgaacccg accgagttcg ggtgatccgc 25080
gaggggtgtg cgtcgtattt gcccgacatt gatcccgagg agacctcgga gtggctggag 25140
tcctttgaca cgctgctgca acgctgcggc ccgtcgcggg cccgctacct gatgttcgg 25200
ctgctagagc gggccggcga gcagcgggtg gccatcccg cattgacgtc taccgactat 25260
gtcaacacca tcccgaaccg gctggagccg tggttccccg gcgacgaaga cgtcgaaagt 25320
cgttatcgag cgtggatcag atggaatgcg gccatcatgg tgcaccgtgc gcaacgaccg 25380

gggtgtgggcg tgggtggcca tatctogacc tacgcgtcgt ccgcggcgct ctatgaggtc 25440
ggtttcaacc acttcttccg cggcaagtcg cacccgggcg gcggcgatca ggtgttcac 25500
cagggccacg cttccccggg aatctacgcg cgcgccttcc tcgaagggcg gttgaccgcc 25560
gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcggggtt gccgtcctat 25620
ccgcaccgc ggctcatgcc cgacttctgg gaattcccca ccgtgtcgat gggtttgggc 25680
ccgctcaacg ccatctacca ggcacggttc aaccactatc tgcattgaccg cggatatcaa 25740
gacacctccg atcaaacagt gtggtgtttt ttgggcgacg gcgagatgga cgaacccgag 25800
agccgtgggc tggccacagt cggcgcgctg gaaggcttg acaacttgac cttcgtgatc 25860
aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920
ctggagtcgt tcttcgcgcg tgccggctgg aacgtcatca aggtggtgtg gggccgcgaa 25980
tgggatgcc tgctgcacgc cgaccgcgac ggtgcgctgg tgaatttaat gaatacaaca 26040
cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccacttc 26100
ttcgcccgcg acccaacgac caaggcgctg gtggagaaca tgagcgacca ggatatctgg 26160
aacctcaaac ggggcggcca cgattaccgc aaggtttacg ccgcctaccg cgccgcgctc 26220
gaccacaagg gacagccgac ggtgatcctg gccaaagca tcaaaggcta cgcgctgggc 26280
aagcatttcg aaggacgcaa tgccaccac cagatgaaaa aactgaccct ggaagacctt 26340
aaggagtttc gtgacacgca gcggattccg gtcagcgacg ccagcttga agagaatccg 26400
tacctgccgc cctactacca ccccgccctc aacgcccccg agattcgta catgctcgac 26460
cggcgccggg ccctcggggg ctttgttccc gagcgagga ccaagtcaa agcgctgacc 26520
ctgccgggtc gcgacatcta cgcgccgtg aaaaagggt ctgggcacca ggaggtggcc 26580
accaccatgg cgacggtgcg cacgttcaa gaagtgttc gcgacaagca gatcgggccg 26640
cggatagtcc cgatcattcc cgacgaggcc cgcaccttcg ggatggactc ctggttcccg 26700
tcgctaaaga tctataaccg caatggccag ctgtataccg cggttgacgc cgacctgatg 26760
ctggcctaca aggagagcga agtcgggcag atcctgcacg agggcatcaa cgaagccggg 26820
tcggtgggct cgttcacgc ggccggcacc tcgtatgcga cgcacaacga accgatgatc 26880
cccatattaca tcttctactc gatgttcggc ttccagcgca ccggcgatag cttctgggcc 26940
gcggccgacc agatggctcg agggttcgtg ctcggggcca ccgccccg caccacctg 27000
accggtgagg gcctgcaaca cgccgacggt cactcgttcg tgctggccgc caccaaccg 27060
gcggtggttg cctacgaccc ggccttcgcc tacgaaatcg cctacatcgt ggaaagcgga 27120
ctggccagga tgtcgggga gaaccggag aacatcttct tctacatcac cgtctacaac 27180
gagccgtacg tgcagccgc ggagccggag aacttcgatc ccgagggcg gctcgggggt 27240
atctaccgct atcacgcggc caccgagcaa cgcaccaaca aggcgcagat cctggcctcc 27300

ggggtagcga tgcccgcggc gctgcgggca gcacagatgc tggccgccga gtgggatgtc 27360
gccgccgacg tgtggtcggt gaccagttgg ggcgagctaa accgcgacgg ggtggccatc 27420
gagaccgaga agctccgcca ccccgatcgg ccggcgggcg tgccctacgt gacgagagcg 27480
ctggagaatg ctcgggggcc ggtgatcgcg gtgtcggact ggatgcgcgc ggtccccgag 27540
cagatccgac cgtgggtgcc gggcacatac ctacagttgg gcaccgacgg gttcggcttt 27600
tccgacactc ggcccgcggc tcgcccgtac ttcaacaccg acgccgaatc ccagggtggtc 27660
gcggtttttg aggcgtttgg gggcgacggc gagatcgacc catcggtgcc ggtcgcggcc 27720
gccgcgcagt accggatcga cgacgtggcg gctgcgcccg agcagaccac ggatcccggc 27780
cccggggcct aacgccggcg agccgaccgc ctttggccga atcttccaga aatctggcgt 27840
agcttttagg agtgaacgac aatcagttgg ctccagttgc ccgccgagg tcgccgctcg 27900
aactgctgga cactgtgccc gattcgctgc tgcggcggtt gaagcagtag tcgggcgggc 27960
tgcccaaccga ggcagtttcg gccatgcaag aacggttgcc gttcttcgcc gacctagaag 28020
cgtcccagcg cgccagcgtg gcgctggtgg tgcagacggc cgtggtcaac ttcgtcgaat 28080
ggatgcacga ccgcacagat gacgtcggct ataccgcga ggcattcgag ctggtgcccc 28140
aggatctgac gcgacggatc gcgctgcgcc agaccgtgga catggtgcgg gtcaccatgg 28200
agttcttoga agaagtcgtg cccctgctcg cccgttccga agagcagttg accgccctca 28260
cgggtggcoat tttgaaatac agccgcgacc tggcattcac cgccgccacg gcctacgcgg 28320
atgcggccga ggcacgaggg acctgggaca gccggatgga ggccagcgtg gtggacgcgg 28380
tggtacgcgg cgacaccggt cccgagctgc tgtccggggc ggccgcgctg aattgggaca 28440
ccaccgcgcc ggcgaccgta ctggtgggaa ctccggcgcc cgggtccaaat ggctccaaca 28500
gcgacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cgccacggcc 28560
gcgctgcgct gaccgacgtg caccggcacct ggctggtggc gatcgtctcc ggccagctgt 28620
cgccaaccga gaagtctctc aaagacctgc tggcagcatt cgccgacgcc ccggtggtca 28680
tcggcccccac ggcgcccatg ctgaccgcgg cgcaccgcag cgctagcgag gcgatctccg 28740
ggatgaacgc cgtcgccggc tggcgcgag cgccgcggcc cgtgctggct agggaaacttt 28800
tgccgaacg cgccctgatg ggcgacgcct cggcgatcgt ggccctgcat accgacgtga 28860
tgcgggccct agccgatgcc ggaccgacgc tcatcgagac gctagacgca tatctggatt 28920
gtggcgggcg gattgaagct tgtgccagaa agttgttcgt tcatccaaac acagtgcggc 28980
accggctcaa gcgcatcacc gacttcaccg ggcgcgatcc caccagcca cgcgatgcct 29040
atgtccttcg ggtggcgggc accgtgggtc aactcaacta tccgacgccg cactgaagca 29100
tcgacagcaa tgccgtgtca tagattccct cgccggctcag aggggggtcca gcagggggcc 29160
cggaagata ccagggggcg cgtcggacgg aaagtgatcc agacaacagg tcgcgggacg 29220

atctcaaaaa catagcttac aggcccgttt tgttggttat atacaaaaac ctaagacgag 29280
gttcataatc tgttacacog cgcaaaaccg tcttcacagt gttctcttag acacgtgatt 29340
gcgttgctcg caccocggaca gggttcgcaa accgagggaa tgttgctgcc gtggcttcag 29400
ctgcccgcg cagcggacca gatcgcgcg tggtcgaaag ccgctgatct agatcttgcc 29460
cggctgggca ccaccgcctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg 29520
atcgtcgocg cgactctgct ggcccaccag gaactggcgc gccgatgcgt gctcgccggc 29580
aaggacgtca tcgtggcccg cactccgctc ggcgaaatcg cggcctacgc aatcgccggt 29640
gtgatagccg ccgacgacgc cgtcgcgctg gccgccaccc gcggcgccga gatggccaag 29700
gcctgcgcca ccgagccgac cggcatgtct gcggtgctcg gcggcgacga gaccgagggtg 29760
ctgagtcgcc tcgagcagct cgacttggtc ccggcaaacc gcaacgcgc cgccagatc 29820
gtcgctgccg gccggtgac cggcttgag aagctcgccg aagaccgccc ggccaaggcg 29880
cgggtgogtg cactgggtgt cgccggagcg ttccacaccg agttcatggc gcccgcaactt 29940
gacggctttg cggcgccgc ggccaacatc gcaaccgccc accccaccgc cacgctgctg 30000
tccaaccgcg acgggaagcc ggtgacatcc gcggccgccc cgatggacac cctggtctcc 30060
cagctcacc caccgggtg atgggacctg tgcaccgcca cgctgcgcca acacacagtc 30120
acggcgatcg tggagttccc ccccgcgggc acgcttagcg gtatcgcca acgcgaactt 30180
cgggggggtc cggcacgccc cgtcaagtca cccgcagacc tggacgagct ggcaaacct 30240
taaccgcgga ctcgccaga acaaccacat acccgctagt tcgatttgta cacaacatat 30300
tacgaaggga agcatgctgt gcctgtcact caggaagaaa tcattgccgg tatcgccgag 30360
atcatcgaag aggtaccggt tatcgagccg tccgagatca ccccgagaa gtcgttcgtc 30420
gacgacctgg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30480
tacggcgta agatccccga cgaggacctc gccggtctgc gtaccgtcgg tgacgttgtc 30540
gcctacatcc agaagctcga ggaagaaaac ccgaggcgcg ctcaggcggt gcgcgcgaag 30600
attgagtcgg agaaccgga tgccgttgcc aacgttcagg cgaggcttga ggccgagtc 30660
aagtgaagtc gccttcacc gctaattggc gtttccccag cgttggtgtg accgcccgtc 30720
cagcgacgac gtcgatctcg ccggacatcg agagcacgtg gaagggtctg ttggccggcg 30780
agagcgcat ccacgcactc gaagacgagt tcgtcaccaa gtgggatcta gcggtcaaga 30840
tcggcggtca cctcaaggat ccggtcgaca gccacatggg ccgactcgac atcgacgca 30900
tgtcgtaagt ccagcggtg ggcaagttgc tggcgggaca gctatgggag tccgccggca 30960
gcccgagggt cgatccagac cggttcgccc ttgttgctcg caccggtcta ggtggagccg 31020
agaggattgt cgagagctac gacctgatga atgcggcgcc ccccggaag gtgtccccgc 31080
tgcccggtca gatgatcatg cccaacggtg ccgcgccggt gatcggtctg cagcttgggg 31140

cccgcgcgg ggtgatgacc ccggtgtcgg cctgttcgtc gggctcggaa gcgatcgccc 31200
acgcgtggcg tcagatcgtg atgggcgacg ccgacgtcgc cgtctcggc ggtgtcgaag 31260
gacccatcga ggcgtgccc atcgcggcgt tctccatgat gcgggccatg tcgacccgca 31320
acgacgagcc tgagcgggcc tcccgccgt togacaagga ccgacgagcc tttgtgttcg 31380
gcgaggccgg tgcgtgatg ctcatcgaga cggaggagca cgccaaagcc cgtggcgcca 31440
agccgttggc ccgattgtg ggtgccggta tcacctcgga cgcctttcat atggtggcgc 31500
ccgcggccga tgggtgttcgt gccggtaggg cgatgactcg ctgctggag ctggccgggt 31560
tgtcgcggc ggacatcgac cactcaacg cgcacggcac ggcgacgcct atcggcgacg 31620
ccgcggagcc caacgccatc cgcgtcgccg gttgtgatca ggccgcggtg tacgcgccga 31680
agtctgcgt gggccactcg atcggcgcg tgggtgcgt cgagtcggtg ctacgggtgc 31740
tgacgtcgc cgacggcgtc atcccgccga ccctgaacta cgagacaccc gatcccgaga 31800
tcgacctga cgtcgtcgcc ggcgaaccgc gctatggcga ttaccgctac gcagtcaaca 31860
actcgttcgg gttcggcggc cacaatgtgg cgcttgccct cgggcgttac tgaagcaoga 31920
catcgcgggt cgcgaggccc gaggtggggg tcccccgct tgcggggcg agtcggaccg 31980
atatggaagg aacgttcga agaccaatga cggagctggt taccgggaaa gcctttccct 32040
acgtagtcgt caccggcatc gccatgacga ccgcgctcgc gaccgacgcg gagactaagt 32100
ggaagtgtt gctggaccgc caaagcggga tccgtacgt cgatgacca ttcgtcgagg 32160
agttcgacct gccagttcgc atcggcgac atctgcttga ggaattcgac caccagctga 32220
cgcggatcga actgcgccg atgggatacc tgcagcggat gtccaccgtg ctgagccggc 32280
gcctgtggga aaatgccggc taccgcgagg tggacaccaa tcgattgat gtgtccatcg 32340
gcaccggcct gggttcggcc gaggaactgg tottcagtta cgacgatatg cgcgctcgcg 32400
gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gcccaacggg gccgccggc 32460
cggtcgggtt ggaacggcac gccaaaggcg gggtagtgac gccggtatcg gcgtgcgcac 32520
ccggcgccga ggccatcgcc cgtgcgtggc agcagattgt gctgggagag gccgatgccg 32580
ccatctcgcg cggcgtggag accaggatcg aagcgggtgc catcgccggg ttcgtcaga 32640
tgcgcatcgt gatgtccacc aacaacgacg accccgcggg tgcgtgccgc ccattcgaca 32700
gggaccgcga cggctttgtg ttccggcagg gcggcgccct tctgttgatc gagaccgagg 32760
agcacgccaa ggcacgtggc gccaacatcc tggcccggt catgggcgcc agcatcacct 32820
ccgatggctt ccacatggtg gcccgggacc ccaacgggga acgcgcggg catgcgatta 32880
cgcgggcgat tcagctggcg ggcctcgccc ccggcgacat cgaccacgtc aatgcgcacg 32940
ccaccggcac ccaggtcggc gacctggccc aaggcagggc catcaacaac gccttggggc 33000
gcaaccgacc ggcggtgtac gcccgaagt ctgccctcg ccaactcggg ggcgcggtcg 33060

gcgcgggtcga atcgatcttg acggtgctcg cgttgcgcga tcaggtgatc ccgccgacac 33120
tgaatctggt aaacctcgat cccgagatcg atttggacgt ggtggcgggt gaaccgcgac 33180
cgggcaatta ccggtatgcg atcaataact cgttcggatt cggcggccac aacgtggcaa 33240
tcgccttcgg acggtactaa accccagcgt tacgcgacag gagacctgcg atgacaatca 33300
tggcccccga ggcggttggc gagtcgctcg acccccgcga tccgctgttg cggctgagca 33360
acttcttcga cgacggcagc gtggaattgc tgcacgagcg tgaccgctcc ggagtgtggt 33420
ccgcggcggg caccgtcaac ggtgtgcgca ccatcgcggt ctgcaccgac ggcaccgtga 33480
tgggcggcgc catgggcgtc gaggggtgca cgcacatcgt caacgcctac gacactgcc 33540
tcgaagacca gagtccatc gtgggcatct ggcattcggg tgggtgcccg ctggctgaag 33600
gtgtgcgggc gctgcacgcg gtaggccagg tgttogaagc catgatccgc gcgtccggct 33660
acatcccgca gatctcgggt gtcgtcgggt tcgcgcggcg cggcgcggcc tacggaccgg 33720
cgttgaccga cgtcgtcgtc atggcgccgg aaagccgggt gttcgtcacc gggcccgacg 33780
tgggtgcgag cgtcaccggc gaggacgtcg acatggcctc gctcgggtgg ccggagaccc 33840
accacaagaa gtccgggggtg tgccacatcg tcgccgacga cgaactcgat gcctacgacc 33900
gtgggcgcgg gttggtcgga ttgttctgcc agcaggggca tttcgatcgc agcaaggccg 33960
aggccgggtga caccgacatc cagcgctgc tgccggaatc ctcgcgacgt gcctacgacg 34020
tgcgctcgat cgtgacggcg atcctcgatg cggacacacc gttcgacgag ttccaggcca 34080
attgggcgcc gtcgatggtg gtcgggctgg gtcggctgtc gggtcgcacg gtgggtgtac 34140
tggccaacaa ccgctacgc ctgggcggct gcctgaactc cgaaagcgca gagaaggcag 34200
cgcgtttcgt gcggctgtgc gacgcgttcg ggattccgct ggtgggtggtg gtcgatgtgc 34260
cgggctatct gcccggtgtc gaccaggagt ggggtggcgt ggtgcgccgt ggcgccaagt 34320
tgctgcacgc gttcggcgag tgcaccgttc cgcgggtcac gctggtcacc cgaaagacct 34380
acggcggggc atacaattcg atgaactccc ggtcgttgaa cgcgaccaag gtgttcgcct 34440
ggccggaagc cgaggtcgcg gtgatgggcg ctaaggcggc cgtcggcatc ctgcacaaga 34500
agaagtggc cgccgctccg gagcacgaac gcgaagcgct gcacgaccag ttggccgccc 34560
agcatgagcg catcgccggc ggggtcgaca gtgcgctgga catcggtgtg gtcgacgaga 34620
agatcgaccc ggcgcatact cgcagcaagc tcaccgaggc gctggcgag gctccggcac 34680
ggcgcggccg ccacaagaac atcccgtgt agttctgacc gcgagcagac gcagaatcgc 34740
acgcgcgagg tccgcgccgt gcgattctgc gtctgctcgc cagttatccc cagcgggtggc 34800
tggtaaacgc gaggcgtcc tcgcatgctc ggacggtgcc taccgacgcg ctaacaattc 34860
tcgagaaggc cggcgggttc gccaccaccg cgcaattgct caggtcatg acccgccaac 34920
agctcgacgt ccaagtgaaa aacggcggcc tcgttcgcgt ttggtacggg gtctacgcg 34980

cacaagagcc ggacctgttg ggccgcttg cggctctcga tgtgttcattg ggggggcacg 35040
ccgtcgcgtg tctgggcacc gccgccgctg tgtatggatt cgacacggaa aacaccgtcg 35100
ctatccatat gctcgatccc ggagtaagga tgcggccac ggctggctctg atggtccacc 35160
aacgcgtcgg tgcccggtc caacgggtgt caggctgtct cgcgaccgcg cccgcatgga 35220
ctgccgtgga ggtcgacga cagttgcgc gcccgcgggc gctggccacc ctcgacgcg 35280
cactacggtc aatgcgctgc gctcgagtg aaattgaaaa cgccgttgct gagcagcgag 35340
gccgccgagg catcgctcg gcgcgcgaac tcttaccctt cgccgacgga cgcgcggaat 35400
cggccatgga gagcgaggct cggctcgtca tgatcgacca cgggctgccg ttgccgaac 35460
ttcaataccc gatacacggc caccggtgtg aaatgtggcg agtcgacttc gcctggcccg 35520
acatgcgtct cgcggccgaa tacgaaagca tcgagtggca cgcgggaccg gcggagatgc 35580
tgcgcgacaa gacacgctgg gccaagctcc aagagctcgg gtggacgatt gtcccgattg 35640
tcgtcgacga tgtcagacgc gaaccggcc gcctggcggc ccgcatcgcc cgccacctcg 35700
accgcgcgcg tatggccggc tgaccgctgg tgagcagacg cagagtcgca ctgcggcccg 35760
cgcagtgcga ctctgcgtct gctcgcgtc aacggctgag gaactcctta gccacggcga 35820
ctacgcgtc gcgatcccgt ggcaccagac cgatccgggt ccggcggtcg aggatatcgt 35880
ccacatccag cgcacctca tgggtcaccg cgtattcgaa ctccgcccg gtcacgtcga 35940
tgccgtcggc gaccggctcg gtgggcccgt cacatgtggc ggccgacgag acgttgcccg 36000
cctcggcccc gtaccgcgc accagcgact cgggcaatcc ggcccccgat ccgggggccc 36060
gccagggtt cgcgggtgc ccgatcagcg gcaggttgcg agtcggcac ttcgcggctc 36120
gcaggtgtcg cagcgtgatg gcgcgattca gcacatctc tgccatgtag cggattccg 36180
tcagcttgcc gccgaccaca ctgatcacgc ccgacggcga ttcaaaaaca gcgtggtcac 36240
gcgaaacgtc ggccgtgcg ccctggacac cagcacgcgc ggtgtcgatt agcggccgca 36300
atcccgcata ggcaccgatg acatccttg tgccgaccgc cgtccccaat gcggtgttca 36360
ccgtatccag caggaaacgtg atctcttcg aagacggtt tggcacatcg ggaatcgggc 36420
cgggtgcgtc ttctcggtc agcccgagat agatccggcc cagctgctcg ggcattggca 36480
acacgaagcg gttcagctca ccggggatcg gaatggtcag cgcggcagtc ggattggcaa 36540
acgacttcgc gtcaagacc agatgtgtgc cgcggctggg gcgtagcctc agggacgggt 36600
cgatctcacc cgcacacag cccgccgctg tgatgacggc acgcgccgac agcgcgaacg 36660
actgccgggt gcgcgggtcg gtcaactcca ccgaagtgc ggtgacattc gacgcgcca 36720
cgtaagttag gatcgggcg ccgtgctggg ccgcgggtcg cgcgacggcc atgaccagcc 36780
gggctcgtc gatcaattgc ccgtcgtacg cgagcagacc accgtcgagg ccgtcccgc 36840
gaacggtggg agcaatctcc accaccgtg acgccgggat tcggcgcgat cggggcaacg 36900

tcgccgccgg cgtacccgct agcaccgcga aagcgtcgcc ggccaggaaa ccggcacgca 36960
 ccaacgcccg cttggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcattggcat 37020
 gcacgagatg aggagcgttg cgtgtcatca ggattccgcg ttcgacggcg ctgcgccggg 37080
 cgatgcccac gttgccgctg gccagatagc gcagaccgcc gtgcaccaac ttcgagctcc 37140
 agcggctggt gccgaacgcc agatcatgct ttccaccaaa ggccaccgtc agaccgcggg 37200
 tggcagcatc taaggcaatg ccaacaccgg taatgccgcc gcctatcacg atgacgtcga 37260
 gtgcgccacc gtcggccagt gcggtcaggt cggcggagcg acgcgccggg ttgagtgcag 37320
 ccgagtgggg catcagcaca aatatccgtt cagtgcgtgg gtaagttcgg tggccagcgc 37380
 ggcggaatcg aggatcgaat cgacgatgtc cgcggactgg atggtcgact gggcgatcag 37440
 caacaccatg gtcgccagtc gacgagcgtc gccggagcgc aactgcccg accgctgcgc 37500
 cactgtcagc cggggcgcca acccctcgat caggacctgc tggctggtgc cgaggcgtc 37560
 ggtgatgtac accctggcca gtcocgagt catgaccgac atgatcagat cgtcaccocg 37620
 caaccggtcg gccaccgca caatctgctt taccaacgct tcccggtcgt ccccgtcgag 37680
 gggcacctcc cgcagcacgt cggcgatatg gctggtcagc atggacgcca tgatcgaccg 37740
 ggtgtccggc cagcgacggt atacggtcgg gcggtcagc cccgcgcgcc gggcgatctc 37800
 ggcaagtgtc acccggcca cgcgtaatc gacgacgag ctgcgccgtg cccgcaggat 37860
 acgaccaccg gtatccgcgc ggtcattact cattgacagc atgtgtaata ctgtaacgcg 37920
 tgactcaccg cgaggaactc cttccaccga tgaaatggga cgcgtgggga gatcccgccg 37980
 cggccaagcc actttctgat ggcgtccggt cgttctgtaa gcaggttgtg ggcctagcgg 38040
 actcggagca gcccgaaactc gaccccgccg aggtgcagct gcgcccgtcc gccctgtcgg 38100
 gggcagacca 38110

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

<400> 25

gaaaagggtg acaagtccta ttttcaagag aagatgactt ttaacagttt tgaaggatct 60
 aaaacttgtg tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga 120
 ttaaaaactt ttgctaattt tccaagtggg agtcctgttt cagcatcaac actggcacga 180
 gcagggtttc ttatatactg tgaaggagat accgtgcggt gctttagttg tcatgcagct 240
 gtagatagat ggcaatatgg agactcagca gttggaagac acaggaaagt atccccaaat 300

tgcagattta tcaacggctt ttatcttgaa aatagtgcc cgcagtctac aaattctggt 360
 atccagaatg gtcagtacaa agttgaaaac tatctgggaa gcagagatca ttttgcctta 420
 gacaggccat ctgagacaca tgcagactat cttttgagaa ctgggcaggt tgtagatata 480
 tcagacacca tatacccagag gaaccctgcc atgtattgtg aagaagctag attaaagtcc 540
 tttcagaact ggccagacta tgctcaccta accccaagag agttagcaag tgctggactc 600
 tactacacag gtattggtga ccaagtgcag tgcttttgtt gtggtggaaa actgaaaaat 660
 tgggaacott gtgatcgtgc ctggtcagaa cacaggcgac actttcctaa ttgcttcttt 720
 gttttgggoc ggaatcttaa tattcgaagt gaatctgatg ctgtgagttc tgataggaat 780
 ttcccaaatt caacaaatct tccaagaaat ccatccatgg cagattatga agcacggatc 840
 tttacttttg ggacatggat atactcagtt aacaaggagc agcttgcaag agctggattt 900
 tatgcttttag gtgaagggtga taaagtaaag tgctttcact gtggaggagg gctaactgat 960
 tggaagccca gtgaagaccc ttgggaacaa catgctaaat ggtatccagg gtgcaaatat 1020
 ctgttagaac agaagggaca agaatatata aacaatattc atttaactca ttcaactgag 1080
 gagtgtctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc 1140
 atcttccaaa atcctatggt acaagaagct atacgaatgg gggtcagttt caaggacatt 1200
 aagaaaataa tggaggaaaa aattcagata tctgggagca actataaatc acttgagggt 1260
 ctggttgag atctagttaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact 1320
 tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt 1380
 tgcaaaatct gtatggatag aaatattgct atcgtttttg ttccttggtg acatctagtc 1440
 acttgtaaac aatgtgctga agcagttgac aagtgtccca tgtgctacac agtcattact 1500
 ttcaagcaaa aaatttttat gtottaatct aactotatag taggcatggt atgttgttct 1560
 tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat 1620
 tagcatttgc taccaagtag gaaaaaaat gtacatggca gtgttttagt tggcaatata 1680
 atctttgaat ttcttgattt ttcagggtat tagctgtatt atccattttt tttactgtta 1740
 ttttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt 1800
 attcatagta tactgattta atttctaagt gtaagtgaat taatcatctg gattttttat 1860
 tcttttcaga taggcttaac aaatggagct ttctgtatat aaatgtggag attagagtta 1920
 atctcccaa tcacataatt tgttttgtgt gaaaaaggaa taaattgttc catgtggtg 1980
 gaaagataga gattgttttt agaggttggt tgttgtgttt taggattctg tccattttct 2040
 tgtaaaggga taaacacgga cgtgtgcgaa atatgtttgt aaagtgattt gccattgttg 2100
 aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca 2160
 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatagtctg agccagatca 2220

aagtatgtat gttgttaata tgcatagaac gagagatttg gaaagatata caccaaactg 2280
 ttaaattgtgg tttctcttcg gggagggggg gattggggga ggggccccag aggggtttta 2340
 gaggggcctt ttcaactttc acttttttca ttttgttctg ttcggatttt ttataagtat 2400
 gtagaccccg aagggtttta tgggaactaa catcagtaac ctaacccccg tgactatcct 2460
 gtgctcttcc tagggagctg tgttgtttcc caccaccac ccttccctct gaacaaatgc 2520
 ctgagtgtctg gggcactttg 2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26
 agcuccuaua acaaaagucu guugcuugug uuucacauuu uggauuuccu aaauaaugu 60
 ucucuuuuua gaaaaggugg acaaguccua uuuucaagag aag 103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27
 ggauuuccua auauaauguu cucuuuuu 28

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28
 ccgccagatt tgaatcgcg gaccggttg cagaggtggc ggcggcgga tgggtgcccc 60
 gacgttgccc cctgcctggc agccctttct caaggaccac cgcatctcta cattcaagaa 120
 ctggcccttc ttggagggt ggcctgcac cccggagcgg atggccgagg ctggcttcat 180
 ccactgcccc actgagaacg agccagactt ggcccagtgt ttcttctgct tcaaggagct 240
 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg 300

cgctttcctt tctgtcaaga agcagtttga agaattaacc cttggtgaat ttttgaaact 360
 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agāāagaatt 420
 tgaggaaact gcgaagaaag tgcgccgtgc catcgagcag ctggctgcca tggattgagg 480
 cctotggccg gagctgocctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540
 gtgccaccag ccttcctgtg ggccccttag caatgtctta ggaaaggaga tcaacatttt 600
 caaattagat gtttcaactg tgctcctgtt ttgtcttgaa agtggcacca gaggtgcttc 660
 tgocctgtgca gcgggtgctg ctggtaacag tggctgcttc tctctctctc tctctttttt 720
 gggggctcat ttttgctgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag 780
 aaggcagtgt cccttttgct agagctgaca gctttgttcg cgtgggcaga gccttcaca 840
 gtgaatgtgt ctggacctca tgttgttgag gctgtcacag tctgagtgt ggacttggca 900
 ggtgocctgtt gaatctgagc tgcaggttcc ttatctgtca cacctgtgcc tctcagagg 960
 acagtttttt tgttgttggtg tttttttgtt tttttttttt ggtagatgca tgacttgtgt 1020
 gtgatgagag aatggagaca gagtccctgg ctctctact gtttaacaac atggctttct 1080
 tattttgttt gaattgttaa ttcacagaat agcacaaact acaattaaaa ctaagcacia 1140
 agccattcta agtcattggg gaaacggggg gaacttcagg tggatgagga gacagaatag 1200
 agtgatagga agcgtctggc agatactcct tttgccactg ctgtgtgatt agacaggccc 1260
 agtgagccgc ggggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc 1320
 ctttttaaat gacttggctc gatgctgtgg gggactggct gggctgctgc aggcogtgtg 1380
 tctgtcagcc caaccttcac atctgtcacg ttctccacac gggggagaga cgcagtcgc 1440
 ccaggctccc gctttctttg gaggcagcag ctcccgagg gctgaagtct ggcgtaagat 1500
 gatggatttg attcgccctc ctccctgtca tagagctgca ggggtggattg ttacagcttc 1560
 gctggaaacc tctggaggtc atctcggctg ttctgagaa ataaaaagcc tgtcatttc 1619

<210> 29

<211> 27

<212> RNA

<213> Homo sapiens

<400> 29

ggcgucacac cuucggguga agucgcc

27

<210> 30

<211> 27

<212> RNA

<213> Homo sapiens

<400> 30
ggcgucacac cuucggguga agucgcc

27

<210> 31

<211> 12

<212> PRT

<213> Homo sapiens

<400> 31

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro
1 5 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/11757

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/02; G01N 27/26

US CL : 435/6; 536/23.1; 204/451

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/23.1; 204/451

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN, BAST**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6329146 B1 (Crooke et al) 11 December 2001 (11.12.2001), column 40, example 11.	1
Y	US 5,807,682 A (Grossman et al) 15 September 1998 (15.09.1998), column 19, lines 2-18.	1
Y	US 6,355,428 (Schroth et al) 12 March 2002 (12.03.2002), column 8, lines 64-67.	1
Y	US 6,320,040 B1 (Cook et al) 20 November 2001 (20.11.2001), column 11, lines 14-22	1
Y	US 6,391,542 B1 (Anderson et al) 12 May 2002 (12.05.2002), column 36, example 18.	1

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:		
"A" document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"B" earlier application or patent published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed	"Z"	document member of the same patent family

Date of the actual completion of the international search

22 June 2002 (22.06.2002)

Date of mailing of the international search report

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks

Box PCT

Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Jyotsna Venkat

Telephone No. (703) 308-3235

Jyotsna Venkat PH.D
EXAMINER
TECHNOLOGY CENTER 1600